

## Urgent thinking required about development

**A report from the World Bank has highlighted the dangers of a growing knowledge gap between rich and poor nations. The issue needs to be placed at the heart of development aid strategies.**

The philosopher of science Francis Bacon put it succinctly at the dawn of the scientific revolution when he wrote: "knowledge is power". The World Bank put forward the same idea in a slightly different way this week when it devoted its annual report<sup>1</sup> on world development to the central role of knowledge in the development process (see page 529). Its interpretation of the word is somewhat broader than Bacon's; while the latter was speaking primarily of what we would now call 'natural science', the bank uses the term to embrace both technical know-how and knowledge of attributes, ranging from the quality of a worker to the credit-worthiness of a company. But its main theme is surprisingly similar: "knowledge is development".

The message is an important one to remember at a time when the global economic system is facing its severest crisis since the Second World War. Ironically, the bank's conclusion is based directly on those East Asian economies whose current problems have been at the heart of this crisis; comparing their previous growth rates to that of others, such as the Soviet Union, better endowed with natural resources, it suggests that the reason for their previous success was an ability to work not harder but more smartly. Being able to generate useful knowledge, or take advantage of others', has become almost by definition the key to survival in a knowledge-based global economy.

In one sense, that isn't new. Western nations, many taking a lead from the Asian economies, have long accepted the need to focus on the generation of scientific knowledge in a way that such efforts are relevant to underlying social and economic needs (see, for example, page 534). But the central role of knowledge, and particularly scientific knowledge, in economic and social development has yet to be properly integrated into the policies of those global institutions responsible for the welfare of developing nations — or indeed into the policies of such nations themselves.

### Intellectual property

Take, for example, the question of intellectual property. Almost everyone accepts that without proper legal protection for innovative ideas, the private sector would not invest in the research needed to produce them, as it would lack the means to secure a profit on its investment. But writing the international rules in a way that those who benefit from them most are those who are already the most economically powerful is not necessarily in the best interests of those who are largely excluded. Indeed, as the World Bank report points out, the current rules may in fact be hindering the technological development of developing countries. Helping poorer countries make the transition to a knowledge-based economy is a tough challenge when the industrialized world increasingly holds all the cards in the form of patent portfolios on which are based not only individual products but whole swathes of technology.

A similar recognition in the limitation of the market lies behind the recent acknowledgement by the World Bank and other international organizations of the failure of market forces — and with it the

patent system — to encourage the investment needed to develop antimalarial drugs (see *Nature* 395, 417; 1998). Again, this has highlighted the broad need for more innovative approaches to investment in research and development into Third World problems, including in this case the careful application of that anathema of free marketers: public interventionism.

### Sustainable development

A meeting of the G7 countries last weekend emphasized the need to think innovatively about global economics in the light of the current crisis. Less dramatically, perhaps, new approaches are required to the role of science and technology in development. It is not merely a question of pumping technical aid into the poor nations of the world; too often this aid fails to take root, and its impact evaporates with the departure of Western technical experts. Nor is it a question of leaving the choice of technological priorities to the market-place.

For the World Bank itself, 'sustainable capacity building' has now replaced infrastructure projects, such as financing big dams and bridges, as the focus of its development strategy. But there is an urgent need to understand better just what sustainable development requires of international science and technology networks — whether academic, government or private.

Next year sees a number of important meetings at which such issues will be on the table, and at which developing countries will have the opportunity to push their case for better terms adapted to closing rather than widening the knowledge gap. One is the renegotiation of the 1994 Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), which requires members of the World Trade Organization to abide by agreed standards of patent and copyright protection, while the poorest developing countries have been given until 2006 to draw their national laws into line.

A second opportunity will be the World Science Conference being organized by Unesco in Budapest next May. It is important that this conference does not become trapped in an agenda devoted to science as a primarily cultural activity or the purely ethical issues raised by biomedical research, at a time when narrowing the North-South knowledge gap is the most pressing social and economic aspect of science requiring attention.

The World Bank's report has focused a welcome spotlight on the need for international organizations — including the bank itself — to give the issues of technology transfer and the central role of science and education in development the same attention as free trade issues have commanded in the past. Ultimately, what may be needed is a new set of ground-rules for research and innovation between the industrialized and developing world. Such an initiative would be particularly timely, given that foreign investment in developing countries is likely to be one of the major victims of the current world financial crisis. □

1. World Bank *World Development Report 1998/99: Knowledge for Development* (Oxford University Press, 1998).



# World Bank calls for a fairer deal on patents and knowledge

[PARIS] Developing countries which are unhappy with new international treaties tightening intellectual property rights (IPR) received unexpected support this week from the bastion of capitalism, the World Bank.

In *Knowledge for Development*, the 21st report in the bank's annual series on world development, the bank argues that stronger international IPR legislation risks "shifting bargaining power towards the producers of knowledge, and increasing the knowledge gap" between industrialized and developing countries. It called on developing countries to take a more proactive and hardline stance in international IPR negotiations.

Traditionally the World Bank, established at a meeting of world leaders in 1944 as part of efforts to reconstruct the world economy after the Second World War, has focused largely on promoting economic development, for example through increased free trade. The new report identifies "knowledge" as a less tangible, but central, factor for improving health and living standards.

"Knowledge can make the difference between sickness and health, between poverty and wealth," Lyn Squire, director of development economics at the bank, told a press briefing in Paris last week. Internet access alone will have a revolutionary impact on Third World development, Squire predicted. And the report warns that the vertiginous growth in global knowledge creates the risk that poor countries may fall further behind.

The 1994 agreement on trade-related aspects of intellectual property rights (TRIPS) has created a global regime which sets minimum standards for IPR protection. It also sets, for the first time, an international legal mechanism, through the World Trade Organization's dispute settlement procedure, to sanction countries which fail to abide by the legislation.

The report points out the need to strike a balance between providing incentives for generating new knowledge, and creating conditions for its dissemination. A bank study of drug and chemical companies in the United States, Germany and Japan found that more than a quarter felt that IPR protection in developing countries including China, India and Argentina was too weak to allow investment or technology transfer.

But it warns that there is now a risk of excessively strict IPR adversely affecting follow-on innovations that draw on patents, and actually slowing the pace. Of particular concern is the current tendency for patents to cover not just products but broad areas of technology, in particular in biotechnology.

"So many industrial-country firms are acquiring strong IPR positions, often covering fundamental research tools as well as marketable products, that it may prove hard for new firms and researchers to elbow into this new global industry," says the report.

Furthermore, the report argues that IPR is not an appropriate mechanism to stimu-

| Country       | A     | B     | C       | D       |
|---------------|-------|-------|---------|---------|
| United States | 442.0 | 3,732 | 127,500 | 107,964 |
| Japan         | 75.8  | 5,677 | 335,061 | 53,896  |
| France        | 49.9  | 2,537 | 16,140  | 73,626  |
| Brazil        | 4.2   | 165   | 2,757   | 23,040  |
| India         | 0.05  | 151   | 1,545   | 5,021   |
| Colombia      | 1.81  | 39    | 141     | 1,093   |
| Madagascar    | 0.03  | 22    | 21      | 15,802  |

A. Internet hosts per 1,000 inhabitants (July 1997)

B. R&D scientists and engineers per million (1981-1995)

C. Patents filed by residents (1995)

D. Patents filed by non-residents (1995)

Source: World Bank

**Worlds apart: gap between richest and poorest is growing in terms of access to knowledge.**

late research in many areas of health and medicine, such as AIDS or malaria, where, as it points out, the "social returns" of an innovation — to all those benefiting from it — far exceed the returns to investors. Here, it says, public authorities have a responsibility to subsidize research or provide financial incentives to the private sector, as recently proposed for development of antimalarial drugs (see *Nature* 395, 417-8; 1998).

At the same time, the bank points out that such concerns need to be balanced against the advantages of stricter IPR. These include greater access to foreign markets and technology for countries which enforce international IPR standards, compared with those whose lack of legislation deters investors.

The report proposes no solutions, but calls on developing countries "to negotiate internationally for intellectual property right regimes that give adequate consideration to their urgent need to narrow the knowledge gap", while maintaining incentives for knowledge producers to invest in research. They should also keep up to speed on "new issues for negotiation, such as biotechnology and information technology".

In a bid to practice what it preaches, the bank — which agreed to loans of \$28,594 billion in 1998 compared with \$19,147 billion last year, has now set itself the objective of becoming a clearing house for information on development. It plans to make its vast inhouse expertise and archives, such as staff reports on various developmental issues, available on the Internet by 2000.

Developing countries need to be assertive in defending the terms under which companies are given access to their resources, says the report. In 1990, world sales of medicines derived from plants discovered by indigenous peoples amounted to \$43 billion, with hardly any financial return to these groups (see *Nature* 392, 535-540; 1998). **Declan Butler**

## Magnetar blasts its way out of theory

[WASHINGTON] An enormous blast of gamma-rays and X-rays from a distant star arrived at Earth on the night of 27 August, giving scientists their best evidence yet for a 'magnetar' — a neutron star with a magnetic field stronger than any other known object in the Universe (see right). The event temporarily overloaded X-ray detectors on several spacecraft.

Several teams reported their observations of the event at a press conference at the US space agency NASA last week. The flash came from the Soft Gamma Repeater known as SGR 1900+14, which lies 20,000 light years away in the constellation known as Aquila.

At least seven spacecraft, including the NEAR asteroid probe near the orbit of Mars and the Ulysses solar probe just inside Jupiter's orbit, registered the passing wave.

Although magnetars have been the subject of theory since 1992, the first one



Disruptive influence: an artist's impression of a magnetar, with magnetic field lines shown in blue.

was not identified until recently (see *Nature* 393, 235; 1998). Their intense magnetic fields — 800 trillion times the strength of Earth's — cause the star's thin, solid crust to ripple and crack, occasionally releasing great bursts of energy. **Tony Reichhardt**

## Recession hits plans for science growth in New Zealand

[HOBART] Researchers in New Zealand are feeling the effects of their country's economic recession. The halt to a projected increase in funding for basic research means that the success rate of applications to the Marsden Fund, the main source of such support, has remained at about last year's level of eight per cent in the 1998 round.

In the fourth year since it was launched, the fund — named after Sir Ernest Marsden, founding secretary of the Department of Scientific and Industrial Research, which the government closed in 1991 — is giving NZ\$6.9 million (US\$3.4 million) to 63 new projects out of 757 applications.

In the 1996 round, 13 per cent of grant applications were successful. But the rate fell to about eight per cent last year when the government required that the grants cover the full costs of the research, with no funding coming from other sources.

The annual allocation for the fund has remained constant this year at NZ\$22 million (US\$10.8 million), with no allowance for inflation. About one-third was allocated to new projects, most going to support continuing projects from previous rounds.

The government had previously stated that support for the Marsden Fund was projected to grow to 10 per cent of the value of the Public Good Science Fund, which has received NZ\$317 million this year. But this goal was dropped in the May 1998 budget (see *Nature* 393, 198; 1998).

Projects from seven broad disciplinary areas were selected by a panel of experts. The life sciences was the most successful category overall, with 19 grants. Universities dominate the successful institutions, with 49 grants; seven are shared among various Crown Research Institutes, and the remaining seven among private research institutes and individuals.

Research already supported by the fund includes the production of the first Bose-Einstein condensate by physicists at the University of Otago last August, and the elucidation of fish navigation by researchers at the University of Auckland (M. M. Walker *et al.* *Nature* 390, 371–376; 1997).

George Petersen, president of the Academy of the Royal Society of New Zealand, described the Marsden Fund as “a key component” in overall science funding in the country, and expressed his “deep disappointment” at its curtailment.

With the country's economy falling into recession, many commentators consider it unlikely that the government will reverse the downward drift in its support of research in the near future.

Peter Pockley

## Intel co-founder funds new centre for biodiversity ...

[WASHINGTON] Computer tycoon Gordon Moore last week pledged \$35 million to set up a Center for Applied Biodiversity Science in Washington, the largest private gift ever made to biodiversity conservation.

Moore, co-founder of Intel Corporation, and his wife Betty will give the money to Conservation International (CI), a nonprofit organization that works in some two dozen countries. Gustavo Fonseca, CI's vice president for Brazil programmes, will be executive director and Harvard biologist E. O. Wilson will chair its advisory council.

“It's not going to be a grant-making facility of the typical kind,” Fonseca says. The centre will bring together experts in science, economics and policy to identify emerging threats to biodiversity around the world. It will create fellowships, hold conferences and form partnerships with existing organizations. “I think we can really leverage a lot of what's being done,” he says.

An example is the \$20 million network of eight biological field stations approved by Brazil's National Research Council, scheduled to begin operating this month. The new centre may participate in data collection at

these sites, similar to the Long-Term Ecological Research (LTER) sites sponsored by the US National Science Foundation. Centre-funded researchers could help develop standardized research protocols for field stations in other countries, says Fonseca.

One advantage the new centre will have over other funding organizations like the World Bank is its ability to disperse money quickly and efficiently, he says.

Among threats to biodiversity, says Fonseca, is “predatory” logging in developing areas like the Amazon basin, where international timber companies are cutting down large expanses of pristine forest. As well as assessing damage, the new centre will help determine policies to mitigate the threat.

Moore chairs the executive committee of Conservation International's board of directors. In announcing his gift last week in San Francisco, he cited CI's “unique ability to make use of available information to implement actions” that slow the loss of biodiversity, and said the new centre will assure that “the world's best minds” are on the job. Fonseca expects the new centre to be staffed and operating within weeks.

Tony Reichhardt

## ... and gives Cambridge a science library



Something for everyone: the library donated by Moore (right) will also house Hawking's archive.

[LONDON] Intel's Gordon Moore (see story above) is to finance the building of a £7.5 million (US\$12.5 million) physical sciences and technology library at the University of Cambridge. It will bring together a number of libraries scattered around the city.

The new library, next to the university's centre for mathematical sciences, will include the papers and archive of Stephen Hawking, which the physicist has promised to donate. Hawking is an enthusiastic user of Intel technology: among other applications he reaches the Internet via a wireless GSM (Global System for Mobile communications) connection and a notebook specially

modified for his text-to-voice synthesis software by Intel engineers.

According to the university, the library “represents a new intellectual focus for Cambridge science and technology”. It has been designed to provide “cutting edge electronic retrieval services”, supported by hard copies of journals where available.

Peter Fox, the university librarian, says the circular building — similar to the library designed by Thomas Jefferson at the University of Virginia — has been made “as flexible as possible”. Shelves can be taken out and terminals added as demand for access to electronic data increases.

INTEL



# Congress smiles on research budgets

[WASHINGTON] Major US science agencies are set to enjoy their healthiest budget increases for years, following Congress's agreement to match or exceed most of the funding requests made for them by President Bill Clinton in February.

Despite the expectation of hectic last-minute negotiations between Congress and the White House to finalize the government's budget for the 1999 financial year, which started on 1 October (see *Nature* 391, 521–522; 1998), the two sides have agreed funding for almost all the major science agencies, apart from the National Institutes of Health (NIH).

Funding for the NIH will probably be caught up in these negotiations. But the biomedical research agency is virtually assured of a significant increase over its 1998 budget of \$13.6 billion. The House of Representatives is calling for 9 per cent budget growth and the Senate for 14 per cent.

Overall, the federal government's support for basic research is poised to grow substantially in real terms for the first time since the early 1990s. This is despite the failure of the Congress to raise new money from tobacco taxes which Clinton said in February would pay for his science initiatives.

The energy and water appropriations bill, which was completed on 25 September and which Clinton is expected to sign, will support small increases for nuclear and high-energy physics at the Department of Energy, and flat funding for fusion science. But virtually none of the additional money for energy supply research requested by the administration for its Climate Change Technology Initiative has been granted by the Congress (see *Nature* 394, 305; 1998).

The energy budget contains only \$130 million of the \$157 million requested to start building the Spallation Neutron Source at Oak Ridge National Laboratory, raising the likelihood that the facility will open late.

Under an appropriations bill for veterans' affairs, housing and urban development and independent agencies (VA-HUD), agreed by conference between the House and Senate last week, the National Science Foundation will get \$3.67 billion, a rise of 7 per cent.

The foundation's research account increases by 9 per cent — slightly better than the amount proposed by the Senate, but less than the 12 per cent rise the White House had requested. Following the Senate's lead, funding was denied for the Polar Cap Observatory and \$10 million was added for plant genome research. But a proposed 5 per cent rise for salaries and expenses, which the Senate had cut, was restored.

The VA-HUD negotiations eased some of NASA's financial problems with a decision to honour the administration's full request of

\$2.27 billion for the International Space Station — \$170 million more than the House had granted. A House-proposed addition of \$1.6 million for a ground-based asteroid detection programme was trimmed to \$1 million, and a \$20-million Senate addition for advanced technology for space science was halved.

It is still undecided whether Vice-President Al Gore's proposed Triana Earth-viewing satellite will receive funding (see *Nature* 394, 213; 1998). Last week's compromise bill omitted a House clause denying money for the project pending a congressional review of the proposals received by NASA. The space agency has a winning concept in mind, and hoped to gain final approval from Congress this week. NASA's overall budget of \$13.7 billion is \$200 million more than the White House had requested.

The defence appropriations bill includes increases of 6 per cent for both basic and applied research — the first real rise in six years for the Pentagon programmes which support most engineering and computer



Delayed opening? The Spallation Neutron Source is funded, but it could be finished late.

science research at US universities.

Budgets for smaller research programmes at the departments of agriculture, commerce and the interior are, like those at NIH, yet to be finalized. These may end up in a single 'omnibus' spending bill which Congress will try to negotiate with Clinton before members return to their districts for November's elections.

Colin Macilwain & Tony Reichhardt

## Brussels rejects 'grace period' on patents

[PARIS] The European Commission has rejected a proposal that Europe should unilaterally adopt a 'grace period' allowing researchers and inventors to file for a patent up to a year after publishing a discovery.

At present, a patent has to be applied for before publication. This contrasts with the situation in the United States, where a 12-month post-publication grace period exists, and a committee of the European Parliament has recently proposed that a similar measure be introduced in Europe.

But a public meeting on the issue organized in Brussels on Monday (5 October) by the European Commission heard industry's view that such a move would be a backwards step. The industrialists agreed that the idea appears superficially attractive, particularly to scientists who have rushed into print before realizing that this sabotages their future patent rights.

But the industrialists say it would disadvantage small companies and academic researchers, particularly as it would open the door to expensive litigation. Many, including the British Technology Group (BTG) and the Union of Industrial and Employers' Confederations of Europe (UNICE), are concerned that a grace period would encourage more researchers to publish before patenting, ignoring the fact that this would open their claims to greater contention.

Whereas priority is attributed in Europe on a clear 'first to file' basis, the US system is based on 'first to invent'. The legal ambiguities that this can create inevitably lead to

expensive legal squabbles to establish the paternity of inventions. The United States is the only country to have such a system.

The 'first to invent' system "makes lawyers very rich and unlucky inventors very poor," says Sue Scott, technology licensing officer of BTG, arguing that the cost of meeting a legal challenge can easily run into "six figure sums".

Patenting before publishing is always a more prudent strategy, argue BTG and UNICE. They point out, for example, that if European inventors published details of their inventions before patenting these, a third party might publish, and apply for patents, on minor modifications that the original inventors might have intended to patent.

Worse, third parties might republish the original invention within the grace period in a form where the copying was not apparent, in a bid to obtain patents. Whatever the drawbacks of the existing European system, it has the advantage of being clear cut. Introducing a "grace period" would only "complicate the current system," says Jérôme Chauvin, an adviser to UNICE.

BTG agrees that adoption of a grace period by Europe would be a retrograde step for inventors, and should not be taken unless the United States adopts a first-to-file system.

As the result of a hearing, the commission has abandoned the idea of a grace period "in the near future", according to one official, adding that Brussels will direct its efforts to educating inventors better on the pitfalls of patent law.

Declan Butler & David Dickson

# German scientists may escape fraud trial

[MUNICH] Prosecutors in Germany are finding it more difficult than expected to bring legal charges against two scientists alleged to have perpetrated Germany's biggest ever scientific fraud.

More than a year after a scientific investigation committee concluded that clinical researcher Friedhelm Herrmann and his colleague, molecular biologist Marion Brach, had systematically fabricated data in 37 publications — and nearly two years after the affair first came to light (see *Nature* 387, 442; 1997) — no case has been brought to court.

Brach, who worked — and lived — with Herrmann at Harvard, Freiburg and Berlin before parting from him in 1996 to go to the University of Lübeck, has admitted fabricating data. Herrmann, who moved to the University of Ulm in 1996, continues to maintain his innocence, placing full blame on Brach.

Brach was dismissed as full professor in 1997. She is now reported to be working in New York, and is only communicating with investigators through her solicitor. Herrmann resigned his position in Ulm last month and is now working as a private practitioner in Munich.

Public prosecutors in three regions began investigating different aspects of the affair last year. Those for the state of Baden-Württemberg, Herrmann's formal employer at the University of Ulm, had hoped to take disciplinary action against him on the grounds that evidence that he perpetrated fraud made him "unsuitable" for employment as a professor and *Beamter* (civil servant).

But they had to abandon their strategy

three weeks ago when the ministry of science and research in Baden-Württemberg accepted Herrmann's resignation.

Meanwhile, prosecutors in Ulm have been investigating evidence that Herrmann used research papers containing fraudulent data to support his application for his post as professor at the university. But they have suspended their investigations pending clarification of a court ruling in an unrelated case in Berlin.

This case concerns a policeman from east Berlin, who had concealed his former relationship with the Stasi in order to secure employment in the police force in reunified Germany. The court ruled that he should not be disciplined on the grounds that his work since reunification had been satisfactory.

A verdict from Germany's highest court, the *Bundesgerichtshof*, to which this decision has been appealed, is expected shortly. Because the charges against Herrmann relate to work conducted before he arrived in Ulm, the prosecutors there are waiting for this ruling before deciding whether to proceed.

But Albin Eser, director of the Max Planck Institute for International Criminal Law in Freiburg, who has specialized in scientific fraud, argues that the link between the two cases is not obvious, as the concerns about the policeman as a Stasi informer related to his character, while those about Herrmann related to the way he conducted his job.

Scientists who want to see Herrmann and Brach sanctioned are now pinning their hopes on an investigation by public prosecutors in Berlin, where many of the fraudulent research papers were written, into whether



THOMAS KLING

**Herrmann: Continues to deny any involvement in misconduct, blaming his collaborator.**

the two researchers made false statements to acquire grant money.

The Deutsche Forschungsgemeinschaft (DFG) and the Deutsche Krebshilfe Stiftung (German Cancer Aid Foundation), which between them gave around DM3 million (US\$1.8 million) in research grants to the researchers, have provided documents to help the prosecutors decide whether the success of their grant applications was dependent on fraudulent papers.

The Thyssen Foundation, which awarded a grant of DM200,000 to the pair, has also provided documents to help the prosecutors decide whether the application was copied

## Pressure grows for relaxation of French embryo research laws

[PARIS] French laws banning human embryo research should be eased to allow the use of 'surplus' embryos generated by *in vitro* fertilization (IVF) procedures, according to the country's national bioethics committee and National Academy of Medicine. But the research should only take place under carefully controlled conditions.

These are the conclusions of reports solicited from the two bodies by Bernard Kouchner, the junior health minister, in the run-up to a revision later this year of France's bioethics legislation. This was adopted in 1994 with the provision that it should be reviewed within five years to take account of changes in science and in society (see *Nature* 369, 599; 1994).

Current legislation only permits research on a human embryo that does not harm its 'integrity', which in practice puts a ban on embryo research. This contrasts starkly with the position in Britain, where legislation passed in 1990 allows research on embryos

less than 14 days old, and embryos can be created specifically for research purposes (see *Nature* 344, 799; 1990).

As in Britain, much of the opposition to human embryo research in France comes from groups who argue that life begins at fertilization. But there is also considerable support in France for rejecting what is seen as a utilitarian view of human embryos. Axel Kahn, a prominent geneticist and member of the national ethics committee, says that France does not want to go down the British road "where the embryo is a thing until 14 days, and becomes a human after".

The deliberate creation of embryos for research is likewise vigorously rejected by both the academy and the national bioethics committee. The two bodies each recommend limiting any relaxation of the ban to the use of surplus embryos generated by IVF. The national ethics committee adds that such use should require the consent of the parents, while all research projects involving

embryos should be individually approved by a national commission.

The report of the working group of the National Academy of Medicine argues that human embryo research is a medical "duty" in that it is required to improve IVF and reproductive medicine. Indeed, some observers argue that the current French ban is hypocritical, as the country's IVF techniques benefit from embryo research carried out elsewhere.

Both groups argue that embryos used for research *in vitro* should not be reimplanted. Indeed, one loophole in the current French legislation is that, in theory, it would allow research on embryos provided that this was done for the benefit of the fetus and did not interrupt its development.

"In principle, the law allows you to research the effect of drugs and so on with the experimental outcome being the birth of the child; it's terrible," says Kahn, who wants this loophole closed.

from a Dutch research grant application that Herrmann had been asked to referee.

All three agencies are awaiting the outcome of the Berlin investigation and any possible trial before beginning their own steps to reclaim their money. Such a process could take years. If a trial were to prove that the pair were guilty of fraud, says Christoph Schneider, director for scientific and international affairs at the DFG, it would be a straightforward matter to sue for return of money.

Indeed, in the event of conviction, state-supported grant agencies and charities would be legally obliged to sue for return of grant money. But actually getting the money back would be fraught with difficulty, admits Bruno Zimmermann, the DFG section head who followed the case.

As grant agencies make contracts with institutions rather than individual researchers, the universities where the grant money was spent would presumably have to be sued first, he says. The legal responsibilities have yet to be sorted out, says Zimmermann.

Eser says he is "highly frustrated" with how slowly the case is moving, and is also worried that even if Herrmann's involvement is demonstrated, he may "get away without sanctions". Detlev Ganten, director of the Max Delbrück Centre in Berlin at which Herrmann and Brach worked for several years, is similarly angry at this possibility.

But not everyone is seeking retribution. Guido Adler, dean of medicine at Ulm University, points out that Herrmann is no longer working in academia, and believes that the most important issue is the work of a newly created task force, funded by the DFG, which will assess the scientific impact of the affair and set the record straight (see Box).

## Task force set up to determine the damage

[MUNICH] Determined to assess the full extent of any scientific damage inflicted by the Herrmann and Brach affair (see opposite), the Deutsche Forschungsgemeinschaft (DFG), Germany's main university research funding agency, is funding a task force to pick through the details of around 500 publications that could have been affected.

In April, Ulf Rapp, a professor of biology/immunology at the University of Würzburg, was awarded a one-year grant to conduct the investigation. This will go considerably further than the investigation by a national scientific committee set up

jointly last year by the three German institutes where Herrmann and Brach worked.

The national committee had identified 37 papers in which it concluded that data had either certainly or "most probably" been falsified, often by mixing computer-stored images from different experiments to create new figures.

Rapp's team are examining data and figures in all papers published by Herrmann and Brach, and also some published by former colleagues.

They will try to determine which data may have been fabricated or duplicated, and the origin of figures and their

primary data. Co-authors on the papers have been asked for relevant information about how figures were created and who generated the data.

Rapp was selected because his field of research overlaps with that of Herrmann and Brach, and also because he returned to Germany in 1994 after 25 years abroad.

"The DFG wanted someone independent to head the task force, and because of my absence I was not part of a local network," Rapp says. The investigation's results will be presented to the DFG next year and may be used in court proceedings. **A.A.**

Meanwhile Herrmann, who now works in a private medical practice Munich, continues to deny any involvement in misconduct, and sees the failure to bring charges against him as proof of his innocence. He says he is a victim of press harassment which "has harmed my career and destroyed my family".

"Brach was the scientist in the lab: my main and only job in the past ten years has been to care for patients", he says, contradicting the views of many former colleagues that he has always been more of a lab researcher than a clinician. "Now I just want to be left in peace to build up some sort of future."

He could well achieve his goal. The German Chambers of Physicians, which regulate the medical profession, is informed automatically of court convictions — but does not require doctors on their lists to submit details of criminal investigations.

And if Brach is indeed living in New York, she may well avoid a trial because Germany has only an extradition agreement — under which cases are negotiated individually, and the outcome often depends on the severity of the charges — rather than an extradition treaty, in which extradition is automatic on request.

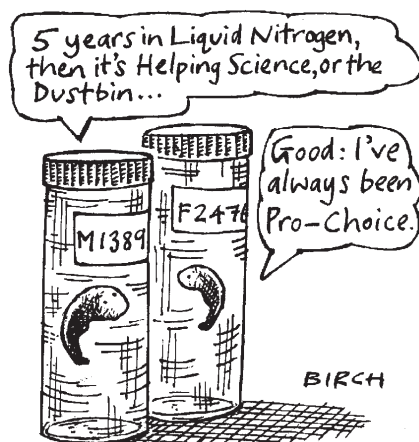
**Alison Abbott**

IVF drugs cause superovulation, resulting in numerous excess oocytes. Since these cannot be stored frozen, but embryos can, the oocytes are fertilized. France has tens of thousands of these excess embryos in storage. At present, the law only allows for couples who have decided not to conserve their surplus embryos to donate them to a sterile couple, or for unwanted embryos to be destroyed after five years in storage.

Critics such as Kahn argue that it is hypocritical to allow the destruction of surplus embryos but to ban their use for research. "They would be destroyed in any case," he says.

A spokesperson for Jean-François Mattei — a physician who opposes human embryo research and, as a member of parliament (Démo­cratie Libérale, Bouches-du-Rhône), was the main architect of France's bioethics legislation — agrees that the fate of surplus embryos has raised difficult questions.

But he argues that recent technical



progress in freezing oocytes means that surplus embryos will soon no longer be produced. Kahn argues that this is still "far off", and that the committee's recommendations only relate to the contemporary situation.

Pressure for relaxing the embryo

research ban also comes from many scientists within the research agencies. The Centre National de la Recherche Scientifique (CNRS), for example, is discussing with the government how the bioethics laws should be revised to take more account of researchers' other concerns — such as the copious paperwork required for even the most innocuous human genetics research.

Scientists are particularly keen to see the law on embryo research liberalized in order to explore the therapeutic opportunities opened up by progress in cloning and the creation of embryonic stem-cell technologies. The national bioethics committee has already argued that the potential of the latter area is such that it requires a dispensation from the current ban (see *Nature* 387, 218; 1997).

The current Socialist government is, in principle, likely to be more sympathetic to researchers' needs than its conservative predecessors.

**Declan Butler**



# Shake-up on the way for Swiss research

[MUNICH] Switzerland's main grant agency for basic research, the National Science Foundation (NSF), is seeking to increase the role of university researchers in identifying the topics of national research programmes — an idea that appears to have the government's support.

At the same time, however, the government is exploring new ways of coordinating and evaluating the teaching and research activities of universities, a move that some of these institutions feel could undermine their newly won autonomy.

At present, the Swiss government decides which targeted applied research programmes — or *Schwerpunktprogramme* — to finance. Seven such programmes are currently being funded on topics such as the environment, biotechnology, and information and communication, at a total cost of SFr63 million (US\$46 million).

The NSF administers the programmes and provides 60 per cent of their budgets. The remaining money comes from the joint council of Switzerland's two technical universities (ETHs) in Zürich and Lausanne.

But the NSF is critical of the top-down decision-making approach in the *Schwerpunktprogramme*. Hans Peter Hertig, the foundation's secretary general, says that they are poorly integrated with other research at the universities in which they are based.

This is because the universities do not identify with research projects they have not been involved in selecting.

The NSF wants the *Schwerpunktprogramme* replaced by a new programme called 'national research priorities', which would give universities more freedom to determine individual research projects. The Swiss Department of the Interior (EDI) supports this idea, and will propose that the government should fund it, in its draft research and higher-education funding scheme for



Swiss universities, such as Lausanne (above), look set to win a greater say on applied research projects.

2000–2003, to be published in November.

According to the proposal, 15 national research priorities — long-term interdisciplinary projects beyond the scope of a single university — would be created by 2003. Each would be modelled on the US National Science Foundation's Science and Technology Centres and similar projects in Germany. They would be based at universities, and involve interdisciplinary collaboration between academic and industrial scientists.

Researchers would be able to propose specific projects falling into one of the five areas of life sciences, environmental and climate research, information technologies and social sciences. Previously the choice of projects had been made by the government. The new programme should be announced in January, and, if approved by parliament, research on the first national research priorities should start in spring 2000. Funding would be expected to increase from SFr30 million in 2000 to SFr75 million in 2003.

The EDI would expect the NSF to reserve SFr148 million of its proposed SFr1,280 million 2000–2003 budget for co-funding the national research priorities, with the rest of

the money coming from the ETH council and private foundations. But the NSF says it will do this only if its total budget is increased, as it does not wish to eat into funds earmarked for basic research.

The NSF is confident that the government will eventually make more money available. But Beat Vonlanthen, vice-director for science and research at the EDI, says that "due to the government's precarious financial situation" no increase could be considered before at least 2003. The agency's budget has decreased over the past four years, a fact that many scientists fear could threaten the traditional high standard of Swiss research (see *Nature* **388**, 817; 1997).

Financial constraints have already led the EDI to propose streamlining university funding. Swiss universities, which are run by the cantons, as well as the federally run ETHs, have won significant financial and scientific autonomy in the past five years.

But the EDI now wants to re-integrate parts of the higher education and research system by creating a central body called *Universitätskonferenz*, to which the cantons would delegate responsibility for funding decisions. The *Universitätskonferenz* would also make the final decisions on which national research priorities are funded.

"We need to delegate policy and decision-making to a central body to improve collaboration between universities and federal research institutes and the coherence of university education," says Vonlanthen.

The EDI also wants to centralize the evaluation of teaching and research, and plans to establish a central institute for university quality control in Bern.

But university rectors have mixed feelings about the proposed changes, which they feel would undermine their autonomy. Georges Fischer, secretary general of the Swiss Conference of University Rectors, says he feels uneasy. "We could be in for a conflict between promised autonomy and new orders from above." Quirin Schiermeier

## Canada's blood agency up and running

[MONTREAL] The Canadian Blood Services agency has started operation, replacing the Canadian Red Cross as the country's national blood service, and it has committed 10 per cent of its Can\$350 million (US\$228 million) budget to research.

The emphasis of the research will be on safety and the reduction of the use of blood, and the government has pledged Can\$125 million over five years to improving the safety of the blood supply system.

The changes follow an inquiry into the tainted-blood scandal, in which HIV-infected blood was used to treat haemophiliacs (see *Nature* **390**, 432; 1997).

One of the chief findings of the inquiry,

led by Mr Justice Horace Krever, was that no single agency seemed to be in overall charge of the blood service. Lynda Cranston, chief executive officer of the new agency, says its most important feature "is our clear line of accountability... the buck stops at CBS".

Judge Krever also called for a no-fault insurance scheme for the agency. But Cranston says this must wait until the provinces come up with enough money for it. A subsidiary, Canadian Blood Services Insurance, will provide cover based on findings of negligence, and an arbitration system will be established to prevent victims of tainted blood transfusions from having to go to court.

David Spurgeon



# Big increase in Spanish research funding

[MADRID] The Spanish government plans to increase spending on civilian research and development by between 8 and 10 per cent next year. The precise figure is uncertain due to its inclusion in a new budget line that also includes military research.

Partly as a result of the inclusion of military projects, such as the development of weapons systems, Spain's total spending on research and development will increase to almost 1 per cent of gross national product (GNP). The country's current research expenditure is among the lowest in Europe.

Recent calculations suggest that the figure fell from 0.91 per cent of GNP in 1991 to 0.76 per cent in 1996, when the comparable figure for neighbouring France was about 2.35 per cent. The increased science funding is part of a broad budget package aimed at preparing Spain for entry into the new European monetary system, which comes into operation next year.

The centre-right government plans to increase total public spending by 4.3 per cent in 1999 over the current year, made possible by its prediction that continued strength in the economy will give a 4.8 per cent increase in revenue, despite cuts in income tax rates.

According to government officials, the aim of the extra spending on research and development is to promote competition between companies and stimulate innovation to banish the widely held view that other countries are better at inventing things.

Plans for research have been outlined within a four-year budget framework. Priorities for investment will include energy research, biotechnology, archaeology, waste removal, and biomedicine and human health. Another priority will be information technology, including improved infrastructure to aid the flow of scientific data.

An Office of Science and Technology will be set up to monitor research spending and report direct to the presidency. Overall, there will be a 9.3 per cent increase in health spending and a 6.5 per cent increase on education.

The extra spending on research and development includes money for developing a range of military technologies, including the production of aircraft, frigates and tanks for the Ministry of Defence, at a cost of US\$1.42 billion.

Although some scientists are critical that this spending is being included in the overall R&D budget, the move is strongly defended

by government officials, who point out that measuring total research effort in this way is a widely accepted practice in other countries.

The extra money for research has been widely welcomed in the scientific community, which has long complained of insufficient funding. Miguel Quintanilla Avila, head of the intracellular signalling department of the Institute of Biomedical Investigations at the Higher Council of Scientific Research (CSIC), says "any state initiative that leads to an increase in public spending on R&D is always welcome".

Quintanilla Avila points out that basic research in Spain will continue to face structural problems, such as the difficulty of obtaining funding for research fellowships, and the high number of researchers on short-term contracts. But he backs the goal of increasing research and development spending to 1 per cent of GNP.

In last week's budget, the government also promises to increase spending on environmental protection by 17.9 per cent, to a total of US\$1.47 billion. Almost four-fifths will be used for water-related infrastructure. The rest will be used for nature conservation and other projects.

**Xavier Bosch**

## French scientist shrugs off winning his second Ig Nobel prize

[BOSTON] French biomedical researcher Jacques Benveniste is set to become the first person to win two 'Ig Nobel' prizes when this year's awards are announced at a ceremony due to take place at Harvard University tonight (8 October).

Benveniste won his first 'Ig' — awarded annually by Marc Abrahams, editor of *The Annals of Improbable Research*, and a group of scientists — in 1991 for his work claiming to show that antibody solutions retain their biological effectiveness, even when diluted to the point where no trace of the antibody is detectable (E. Davenas *et al. Nature* 333, 816–818; 1988). The water, Benveniste argues, preserves a 'memory' of the substance after it is gone.

The second Ig Nobel prize will be awarded for an extension of this work. Benveniste now claims that a solution's biological activity can be digitally recorded, stored on a computer hard drive, sent over the Internet as an attached document and transferred to a different water sample at the receiving end (see [www.digibio.com](http://www.digibio.com)).

"We've demonstrated that you can transmit the biological effect by e-mail between Chicago and Paris," says Benveniste, who heads the Digital Biology Laboratory in Clamart, near Paris, which is financed by the private company DigiBio SA. "With this approach, you could transfer



Unforgettable: Benveniste believes water has a memory and hopes to prove drug effects can be e-mailed.

the activity of a drug by means of standard telecommunications technology."

"French science has not risen to such giddy heights since N-rays were invented by Blondlot early this century," says magician and sceptic James Randi, author of the forthcoming book *A Magician in the Laboratory*, based partly on his involvement in an investigation of Benveniste's laboratory practices carried out in 1988.

Benveniste argues that the science establishment is inherently resistant to new ideas. "Orthodox people are determined to block anything new in biology," he says.

He compares the conventional view that "you need a molecule to have a biological effect" to the debate between Descartes and Newton four centuries ago over whether

action at a distance was possible. "I say the effect comes not from the molecule itself but from the signal it imparts."

Benveniste says he is "happy to receive a second Ig Nobel prize, because it shows that those making the awards don't understand anything. People don't give out Nobel prizes without first trying to find out what the recipients are doing. But the people who give out Ig Nobels don't even bother to inquire about the work."

Harvard chemist Dudley Herschbach, who won the 1986 Nobel Prize in Chemistry, finds Benveniste's claims "hard to reconcile with what we know about molecules". Herschbach considers the prize "very well deserved. And he just might win a third one if he keeps going in this way."

**Steve Nadis**

## US agency formed to tackle threat of 'high-tech' terrorists

[WASHINGTON] The US Department of Defense has formed a Defense Threat Reduction Agency, with 2,100 staff and an annual budget of \$1.9 billion, to be responsible for combating the threat of chemical, biological and nuclear warfare. The agency, which senior Pentagon officials say will expand sharply to deal with what the United States regards as the growing threat of terrorist attack using these weapons, will be led by Jay Davis, a nuclear physicist from the Lawrence Livermore National Laboratory in California.

The agency will be formed by the amalgamation of the Defense Special Weapons Agency, the On-Site Inspection Agency, and the Defense Technology Security Administration. These all focused on issues concerning the nuclear arsenal of the former Soviet Union, but the new agency will direct its attention to a wide range of potential sources of terrorist attack.

Responsibilities of the agency will include co-operative threat reduction programmes, arms control treaty monitoring and site inspections, nuclear, biological and chemical defence, and counterproliferation. An

advisory panel has been established to supervise the agency. Its members include Norm Augustine, former chairman of Lockheed-Martin, John Deutch of the Massachusetts Institute of Technology, and Paul Robinson, president of the Sandia National Laboratory.

## Scientists of tomorrow win awards today

[LONDON] Five young scientists from Austria, Hungary and Britain, aged 15 to 20 years old, are the winners of the 10th European Union contest for young scientists, held last month in Porto, Portugal.

The winners, chosen by an international jury of 14 eminent scientists, succeeded with such projects as a scanning tool for three-dimensional objects, a virtual cane for blind people, and an analysis of the chemical reactions involved in the yellowing process of paints in the dark. The three first prizes are each worth ECU5,000 (US\$6,000).

## India ordered to act over patents dispute

[NEW DELHI] The World Trade Organization (WTO) has set 22 October as the deadline by which India has to introduce either product patents or exclusive marketing rights (EMR)

for pharmaceutical and agricultural chemicals. WTO's dispute settlement board had adopted a report by a panel which investigated the European Union's complaint, lodged last year, that Indian patent laws were not in accordance with the agreement on trade-related aspects of intellectual property rights.

India, which lost a similar case filed by the United States in July 1996, is now left with two choices. It can either introduce product patents right away, or opt for the transition period until 2005 available for EMR for products which have been patented elsewhere. While parliamentary approval is still doubtful, both the industry and the ruling Bharatiya Janata Party are in favour of amending the Indian Patent Act (1972) — if necessary, through a presidential ordinance (see also page 529).

## Japan's nuclear industry gets new public face

[TOKYO] Japan's Nuclear Cycle Development Institute, a semi-governmental organization promoting the nation's nuclear fuel recycling programme, opened last week to replace the Power Reactor and Nuclear Fuel Development Corporation (PNC), which closed on 30 September after 31 years.

The president of the new organization,

Yasumasa Togo, formerly chairman of the government's Nuclear Safety Commission, said at a news conference that the main priority was to restore Japan's trust in nuclear power following a series of nuclear accidents and subsequent cover-ups at PNC facilities last year. The new organization will focus on establishing a nuclear recycling system, disposing of high-level nuclear waste, and developing fast-breeder reactors.

## Los Alamos lab boosts space research work

[WASHINGTON] The Los Alamos National Laboratory in New Mexico has established a Center for Space Science and Exploration to expand its role in space research. The centre will expand an existing office at Los Alamos that oversees about \$9 million in annual funding from the US space agency NASA.

An additional \$5 million will be used to promote space-related technology development and interdisciplinary research in areas such as planetary science, the biological effects of space travel, nuclear power and propulsion systems, and new spacecraft materials.

Los Alamos has provided key instruments for several recent NASA missions, including the Lunar Prospector that detected ice on the Moon.

## Software hitches delay astrophysics launch

[WASHINGTON] The launch of the Advanced X-Ray Astrophysics Facility (AXAF) on the US space shuttle, scheduled for 21 January, may be delayed by three to four weeks owing to additional debugging of flight software. Engineers at the US space agency NASA's Marshall Space Flight Center in Alabama are assessing the software testing schedule and minor electrical problems with the spacecraft this week to determine the extent of the delay.

AXAF, the third of NASA's 'great observatories', was originally scheduled for launch last August, but its builder, TRW, Inc., ran into problems assembling and testing the spacecraft. AXAF is designed to operate for at least five years in Earth orbit.

## British women wise up to science on the web

[LONDON] The UK Association for Women in Science and Engineering (AWiSE) has launched a website to provide information to women interested in an academic or industrial career in science and technology ([www.awise.org](http://www.awise.org)).

AWiSE, founded in 1994, is campaigning to increase women's representation in all areas, and at all levels, of the scientific

community, including education, research, industry, administration and the media. The association points out that, although one third of science, engineering and technology students today are women, they make up a mere three per cent of professors and fellows of the Royal Society.

## French science festival is no laughing business

[PARIS] France's annual science festival kicks off this week, but what has in the past been a weekend event has now been extended to a full week. The science ministry has also changed the festival's name from 'La Science en Fête' to the more sober 'Week of Science', with a shift in emphasis away from the popularization of science to the more serious business of how research works and the issues it raises in society.

As well as the usual open days in French laboratories, this year's event will have national and regional exhibitions on such issues as pollution, urbanisation, food safety, communications technology and Earth observation. Some critics lament that some of the "fun" of previous years seems to have been lost. The publicity pack begins with a dry summary of ministerial policy, which is likely to be of more interest to science policy buffs than lay people.



# Long-lived, but not 'aged'

**Sir** — The growing number of elderly people in society is attracting much interest. Because the matter is of much concern to many people, care must be taken when journals report on ageing matters.

Take, for example, the report of N. Ishii *et al.* on the oxygen-sensitive *mev-1* mutant of the nematode worm (*Nature* **394**, 694–697; 1998). This well-known mutant appears to be defective in mitochondrial electron transport, which could explain its high sensitivity to oxygen poisoning, as shown by its decreased longevity. However, the title of the paper links this result not only to oxidative stress, but also to ageing. Further, the front cover of *Nature* inaccurately states “Ageing: role of oxidative stress”, and the entry on the contents summary page was entitled “Ageing gene”. This choice of title is unfortunate.

First, Ishii *et al.* report longevity, not ageing, data, even though the authors equate a shortening of longevity with

premature ageing. Longevity does not provide the same information as ageing data for individuals; many laboratories rightly use biological markers of ageing, such as those provided by the study of behaviour, rather than relying solely on longevity data. Longevity is a measure of duration, not of content. We all know people of the same longevity, say 90 years old, who were bedridden or who were able to go jogging — knowing the longevity of individuals is not the same as knowing their physiological status or quality of life.

Second, *Nature's* title “Ageing gene” gives the impression that there could be genes governing the way we age. It is well-known that many genes may have effects on various features of the ageing process, but it is not correct to suggest that mendelian genes determine the ageing process. It is well-recognized by gerontologists that genes decreasing longevity are not strong arguments for a genetic determinism of

longevity. Many mutants, due to their negative defects, have a low longevity, and attempts to find mutants where lifespan is increased have failed except in nematodes. Human genetic diseases often decrease lifespan, but it is generally accepted that their study is of little help in understanding normal ageing.

Our purpose is not to criticize the interesting work of Ishii *et al.* But we do wish to draw attention to the need to use the word ‘ageing’ with care when reporting on longevity, and we do wish to ask authors and journals to avoid titles that may misrepresent gerontological research.

**Eric Le Bourg**

**Nadège Minois**

Laboratory of Ethology and Animal Psychology,  
UMR CNRS No. 5550,  
Université Paul Sabatier,  
118 route de Narbonne,  
F-31062 Toulouse cedex 4, France  
e-mail: lebourg@cict.fr

## What's in a name (or a number or a date)?

**Sir** — A great deal of faith is placed in impact factors and citation analyses of published work, not least by institutions being assessed for research quality. These measures assume that the primary data on referenced work are of high quality. But this assumption can be far from the truth.

Take a well-known paper by Marion M. Bradford, entitled “A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding” (*Analytical Biochemistry* **72**, 248–254; 1976). This highly cited paper (well over 10,000 citations) describes a very convenient method for the determination of the amount of protein with reference to a protein standard. It has a single author, and is the only paper published by her in this journal during the 1970s.

Inspection of the BIDS/ISI database reveals that, even when the name of the author and the year are correctly specified, the article has been referred to in 156 different ways, 153 of them giving an incorrect volume number or journal title. In a second search, the date of the article was incorrect in 63 ways. In a third search for which the correct year was specified, the second initial of the author has been omitted (75 errors) or given incorrectly

(15 errors). So there are more than 300 ways in which this article has been referred to incorrectly in published work, amounting to more than 2,000 individual citations.

Apart from raising doubts about the validity of citation exercises, errors in the details of referenced work make it time-consuming for the reader to track down the correct article, and do authors a major disservice by underestimating the impact of their work.

This is by no means an isolated example. Similar exercises on other highly cited papers that describe commonly used biochemical procedures, such as those of Lowry *et al.* (*J. Biol. Chem.* **193**, 265–275; 1951) or Laemmli (*Nature* **227**, 680–685; 1970), show that each of these papers has been referred to in at least 200 incorrect ways.

Who is responsible for checking the accuracy of references in published work? The primary responsibility must lie with authors submitting work for publication. The reviewing and editing process does not seem to be effective in preventing mistakes appearing in the published literature. Should the publishers and the compilers of databases have any responsibility in checking the accuracy of the information? Whatever the answer to this question, it is clear that quality-control procedures are not working satisfactorily.

**Nicholas C. Price**

Department of Biological Sciences,  
University of Stirling, Stirling FK9 4LA, UK  
e-mail: ncp1@stirling.ac.uk

## Each to his own

**Sir** — It is perfectly fine to poll leading scientists about God<sup>1</sup>, but scientists are elected to the US National Academy of Sciences because they are great in one area. This election does not imply “superior knowledge”<sup>2</sup> in other subjects, especially one so personal as a belief in God. One British scientist is quoted<sup>1</sup> as writing: “But I don’t think you can be a real scientist in the deepest sense of the word” and “have religious beliefs”. This is condescension, not insight.

There is a saying in theology that “For those who believe, no proof is necessary. For those who do not believe, no proof is possible”. Some religions claim that divination of religious truths, including faith, represents insights bestowed on a person (rather than earned through some ‘deep’ probing). Great scientists are often fascinating persons. If as a whole they do not believe in God, so be it. But let us not extrapolate the implications of such a census to something about deeper universal truths. (The authors of ref. 1 make no attempt to draw such an inference; they merely report the results.)

**Thomas J. Crowley**

Department of Oceanography,  
Texas A&M University, College Station,  
Texas 77843-3146, USA  
e-mail: tcrowley@ocean.tamu.edu

1. Larson, E. J. & Witham, L. *Nature* **394**, 313 (1998).

2. Leuba, J. H. *The Belief in God and Immortality: A Psychological, Anthropological, and Statistical Study* (Sherman, French & Co, Boston, 1916).

# The fate of the Neanderthals

Paul Mellars

**Between about 30,000 and 40,000 years ago, the Neanderthals in Europe were replaced by populations of behaviourally and biologically modern humans. What happened during that period?**

The Neanderthals have had a chequered scientific career. Berated throughout the first half of the century as distinctly retarded, they were suddenly elevated in the 1960s to having been only marginally less accomplished in biological and behavioural terms than modern *Homo sapiens*<sup>1</sup>. But in the late 1980s they were again relegated to an extinct side-line of human evolution<sup>2,3</sup>, and this was confirmed a couple of years ago<sup>4</sup> by the recovery of surprisingly well-preserved mitochondrial DNA from the original Neanderthal skeleton (from Neanderthal itself). European Neanderthals are thought to have diverged from the lineage that eventually gave rise to modern humans at least half a million years ago. Then, around 30,000–40,000 years ago, the Neanderthals were replaced by modern populations — probably from an ultimately African source<sup>5</sup>.

Now that the evolutionary fate of the Neanderthals has been clinched by DNA evidence, the debate concerns how this process of population replacement came about. In particular, what was the extent of contact between the final Neanderthals and the earliest modern humans<sup>6</sup>? This theme dominated a conference\* held in August to commemorate the discovery of a fossil Neanderthal skull at Forbes Quarry in Gibraltar in 1848 (Fig. 1). Any process of population dispersal and replacement must involve some degree of coexistence and potential interaction between the two populations, but do we measure this in decades or millennia? Did the two groups occupy essentially the same, or sharply separated, territories? And how and why did one population, the Neanderthals, eventually dwindle to the point of extinction?

On one point, at least, there seems to be a consensus — that in the southern part of the Spanish peninsula, roughly to the south of the Ebro valley, the local Neanderthals survived for at least 5,000–10,000 years after the arrival of modern populations in the adjacent parts of northern Spain and the Mediterranean coast. A typical Neanderthal lower jaw from the site of Zafarraya in Andalusia has been dated by two separate methods (radiocarbon and uranium series) to around 30,000 years before present (BP).

Equally typical 'Mousterian' Neanderthal technologies have been dated at sites in both southern Spain and Portugal to a similarly late period (J. Zilhao, Natl Inst. Archaeol., Lisbon; L. Raposo, Natl Mus. Archaeol., Lisbon). By contrast, characteristically modern 'Aurignacian' technologies, almost certainly manufactured by anatomically modern humans, have been dated at several sites in northern Spain to around 38,000–40,000 years BP.

The most likely explanation for the prolonged coexistence of these two populations lies in ecological differences between the northern and southern parts of the Iberian Peninsula. The Ebro Valley seems to act as a sharp ecological boundary for several other species, so the biological and behavioural adaptations of the expanding modern populations were probably not adequate to compete with the local Neanderthals — at least until the modern populations had time to develop their own adaptations to cope with the conditions (A. Currant, Nat. Hist. Mus., London; C. Finlayson, Gibraltar Mus.). Interestingly, despite the prolonged period of coexistence, there is no archaeological evidence that the late Neanderthal populations in southern Spain adopted any behavioural tricks or technological innovations from

their new, more technologically advanced neighbours immediately to the north.

A very different pattern emerges to the north of the Pyrenees. Here, the final Neanderthal populations (as represented, for example, at Saint-Césaire in Aquitaine or at Arcy-sur-Cure in northern Burgundy) started to show remarkably similar behavioural patterns to those of the first anatomically modern populations. The Neanderthals produced simple bone tools, perforated animal teeth and other forms of personal decoration, and scattered powdered red ochre across the floors of their living areas<sup>6,7</sup>. But were these new patterns of behaviour simply copied from contemporaneous modern populations, immediately to the south, or did the Neanderthal populations in this region invent them independently, at almost exactly the same time as the modern human populations were dispersing across Europe?

Many participants at the conference found it hard to accept the hypothesis that this evolution of modern human behaviour was totally independent and coincidental. They pointed to direct dating evidence, which seems to show that the new behavioural patterns among the final Neanderthals emerged only after around 38,000 years BP — that is, only after the modern populations were already established in northern Spain (J.-J. Hublin, Musée de l'Homme, Paris; Fig. 2, overleaf). It was argued that the combination of exceptionally high Neanderthal population densities in western France (reflected in the high densities of archaeological sites in this region), together with the conditions in this area (which were cooler, with more open vegetation than in the Mediterranean areas further to the south), delayed colonization by modern human groups. In any event, evidence from



Figure 1 A head of its time — Neanderthal skull discovered in Gibraltar in 1848. Remains such as this allow us to make reconstructions of what the Neanderthals looked like. That shown overleaf is of a Neanderthal woman based on measurements from a 41,000-year-old skeleton found at Tabun in Israel.

\*Gibraltar and the Neanderthals 1848–1948, Gibraltar Museum, 28–30 August 1998.

both north and south of the Pyrenees shows that population replacement was by no means an overnight event, and seems to have depended on local ecological conditions, as well as on the specific behavioural adaptations of the two groups<sup>8</sup>.

The alternative picture — of an independent, coincidental evolution of typically 'modern' behaviour by the final Neanderthal populations, considerably before modern humans arrived in western Europe — would have dramatic implications. We would have to admit that the behavioural and cognitive capacities of the Neanderthals were every bit as advanced as those of modern humans. Not only that, but also that technological, social and cognitive 'modernity' could erupt at different times and places, presumably in response to some as-yet-unidentified cause. The sceptics in the audience pointed to the statistical improbability of this kind of occurrence, and to the conflicts between this picture and the rapidly accumulating dating evidence (O. Bar-Yosef, Harvard Univ.; C. Gamble, Southampton Univ.).

Aside from these debates, studies of Neanderthal skeletal remains reinforce the conclusion that they were a divergent lineage, which probably made no contribution to the evolution of anatomically modern humans.

A range of anatomical features is unique to the Eurasian Neanderthals, from the detailed form of the lower jaw to the complex morphology of the inner ear. These must have evolved after the Neanderthals separated from the ancestors of modern humans

(Y. Rak, Tel Aviv Univ.; M. Ponce de León, C. Zollikofer & R. Martin, Zurich-Irchel Univ.). Distinctively

Neanderthal features have been identified

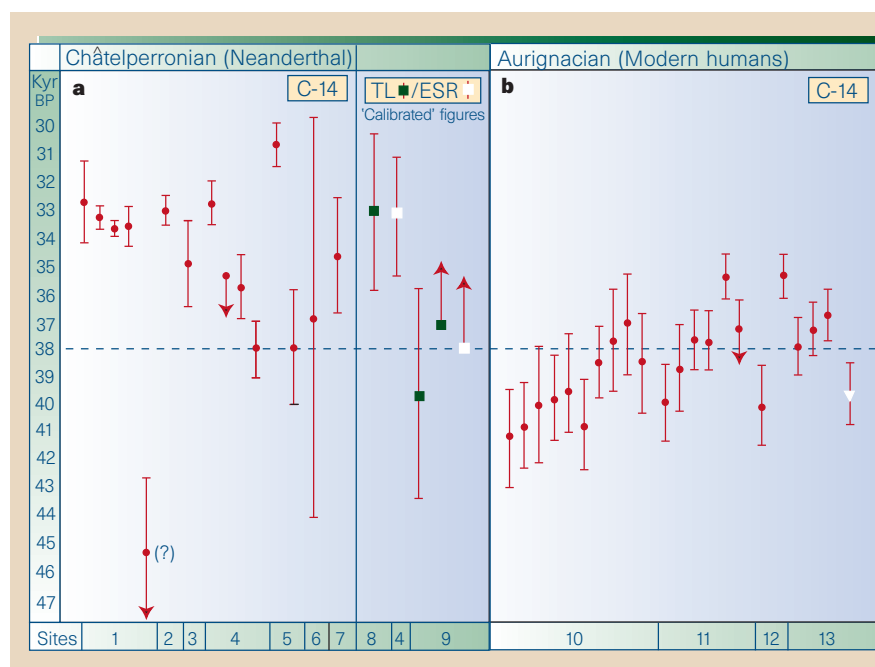
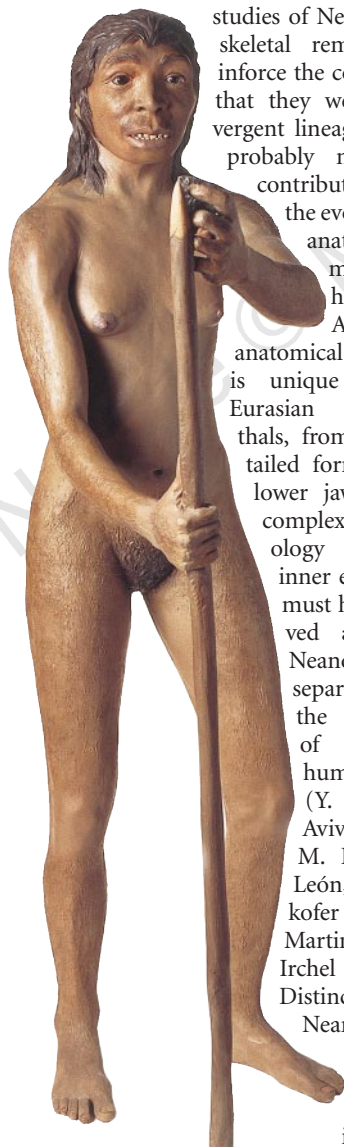


Figure 2 Radiocarbon (C-14) and other absolute age measurements for late Neanderthal and early modern human sites. a, Late Neanderthal, 'Châtelperronian' sites in France. b, Early modern human 'Aurignacian' sites in northern Spain. Note the marked overlap in the age ranges<sup>10</sup>, and that the thermoluminescence (TL) and electron spin resonance (ESR) dates have been 'calibrated' by subtracting 3,000 years, to allow for the known offset between radiocarbon and other dating methods in this time range<sup>11,12</sup>.

in the 35 individuals from the Sima de los Huesos (Atapuerca) in northern Spain, at least 300,000 years BP, and in even earlier skeletons from Petralona in Greece and Arago in southern France (J.-L. Arsuaga, Complutense Univ., Madrid).

All of this is consistent with the DNA evidence that the two lineages separated at least half a million years ago<sup>4</sup>, and even longer divergence times are favoured by some workers. But how much of this divergent development can be attributed to climatic contrasts between Eurasia and Africa? Only the relatively short, stocky bodies of the Neanderthals stand out as a classically climatic adaptation, yet much of their ecological range was in relatively temperate areas of southern Eurasia, as opposed to the periglacial north. Their notoriously large noses and faces were probably due to the pressures of heavy chewing and use of the jaws as tools, rather than to any climatic adaptation to warming up cold air streams in the nasal passages. Most of the other 'robust' features of the Neanderthal skeleton seem to be related directly to their massive body weight, although the morphology of the upper limb bones does seem to reflect exceptionally heavy stresses in the use of the arms and hands (E. Trinkaus, Washington Univ., St Louis).

At the end of the conference we were left with the impression that the Neanderthals really were very different — well adapted to survive in the harsh glacial environments of Europe, perhaps, but with distinct

anatomical and behavioural patterns from their modern human successors. Whether these contrasts reflect different mental abilities remains the 64,000 dollar question. But it could be argued that if the Neanderthals followed a separate evolutionary trajectory from the one that led to modern humans over at least half a million years, it would hardly be surprising if related divergences in cognitive capacities (such as language or intelligence) had evolved over this time<sup>9</sup>. The eagerness of some scientists to claim close kinship with the Neanderthals could come close to denying that human evolution actually took place. □

Paul Mellars is in the Department of Archaeology, University of Cambridge, Downing Street, Cambridge CB2 3DZ, UK.

1. Trinkaus, E. & Shipman, P. *The Neanderthals* (Cape, London, 1993).
2. Cann, R. L., Stoneking, M. & Wilson, A. C. *Nature* **235**, 31–36 (1987).
3. Stoneking, M. & Cann, R. L. in *The Human Revolution* (eds Mellars, P. & Stringer, C.) 17–30 (Edinburgh Univ. Press, 1989).
4. Krings, M. et al. *Cell* **90**, 19–30 (1997).
5. Stringer, C. B. & Mackie, R. *African Exodus: The Origins of Modern Humanity* (Cape, London, 1996).
6. d'Errico, F., Zilhao, J., Julien, M., Baffier, D. & Pelegrin, J. *Curr. Anthropol.* **39**, S1–S44 (1998).
7. Bahn, P. G. *Nature* **394**, 719–720 (1998).
8. Mellars, P. A. in *Neanderthals and Modern Humans in Western Asia* (eds Akazawa, T., Aoki, K. & Bar-Yosef, O.) 493–507 (Plenum, New York, 1998).
9. Mithen, S. *The Prehistory of the Mind* (Thames & Hudson, London, 1996).
10. Mellars, P. *Curr. Anthropol.* (in the press).
11. Bard, E., Hamelin, B., Fairbanks, R. G. & Zindler, A. *Nature* **354**, 405–410 (1990).
12. Laj, C., Mazaud, A. & Duplessy, J.-C. *Geophys. Res. Lett.* **23**, 2045–2048 (1996).



## Planetary science

## Titan weather report

F. Michael Flasar

Cloudy but changeable. That is the outlook for Saturn's huge moon Titan, the only moon in the Solar System to have a thick atmosphere. The workings of that atmosphere are still poorly understood, but on page 575 of this issue<sup>1</sup>, Griffith *et al.* report direct evidence of variable cloud cover.

Our attempts to learn about Titan, particularly its lower atmosphere and surface, have proceeded by fits and starts. When the Voyager 1 spacecraft approached Saturn in 1980 and made the first spatially resolved observations of Titan, its cameras showed a body shrouded in an opaque, brownish haze that blocked the view of the surface at visible wavelengths. Information about Titan's surface and lower atmosphere was limited to spectra in the middle infrared, where a window near 18.5  $\mu\text{m}$  allowed the surface temperature to be determined, and radio-occultation soundings, which yielded vertical profiles of temperature and pressure.

In the occultation experiment, radio waves emitted towards Earth by the spacecraft were refracted as they passed through Titan's atmosphere, giving rise to Doppler shifts in frequency that were used to retrieve a vertical profile of the atmospheric refractivity. For an assumed vertical profile of composition, a vertical profile of temperature and pressure on Titan could be obtained<sup>2</sup>. Combined with ultraviolet and infrared spectra, these radio occultation soundings showed that the mean molecular weight of Titan's atmosphere is about 28 atomic mass units and that most of it is molecular nitrogen, as on Earth. The retrieved surface pressure, ~1.4 bar, is also Earth-like; but the surface temperature, 94 K, is not,

because of Saturn's greater distance from the Sun<sup>3</sup>.

Methane is also known to be an important component of Titan's atmosphere. But it should be destroyed by photolysis in less than ten million years<sup>4,5</sup> — and irreversibly so, because the light hydrogen atoms and molecules that are created escape. What source replenishes it? A global methane ocean was an early favourite after the Voyager encounters, and researchers imagined a 'hydrological' cycle on Titan, with clouds and precipitation not unlike Earth's, but with methane being the condensable gas.

Although the idea of a methane cycle is still reasonable, the hypothesis of a methane ocean quickly ran aground because it is inconsistent with the radio occultation results. One must either assume near saturation of methane just above the surface, implying atmospheric temperatures that decrease more rapidly with altitude than the dry adiabat (and hence are dynamically unstable) in the lowest few kilometres; or a subsaturated profile near the surface, implying fluxes of latent heat (from evaporation) into the base of the atmosphere that exceed by an order of magnitude the solar radiation absorbed at the surface<sup>6</sup>. Neither is physically plausible.

But a pure methane ocean is unrealistic for other reasons. Methane photolysis produces ethane, propane and acetylene in the upper atmosphere, which diffuse downward into the colder stratosphere and tropopause, where they condense and precipitate. Ethane and propane are soluble in methane, and over geological time an initially pure methane ocean would have evolved into a more complex hydrocarbon mix including

dissolved molecular nitrogen<sup>7</sup>. The partial pressure of methane over such an ocean of liquefied natural gas would be lower than that over a pure methane ocean, and consistent with the radio occultation data.

Later observations cast doubt on this idea too. Radar echoes from Titan are too strong to be explained by a smooth surface of liquefied natural gas<sup>8</sup>; the reflectivity is more consistent with a surface of water ice whose deep cracks duct the incident radar signal and reflect most of the energy directly back to the observer. That mechanism is thought to explain the high radar reflectivity of the surfaces of the Galilean satellites of Jupiter. Furthermore, there appear to be persistent features in the near infrared (Fig. 1), implying a solid surface<sup>9,10</sup>. So any global ocean seems unlikely. An ocean/landmass system similar to Earth's is also implausible, because the tidal dissipation caused by such a configuration would circularize Titan's eccentric orbit about Saturn<sup>11</sup>. But lakes of liquefied natural gas are still possible.

Clouds would be the other visible feature of a methane cycle, but they have been elusive. Initially, estimated absorption cross-sections of nitrogen and methane gases were too small to account for the low brightness seen by Voyager<sup>12</sup> at 50  $\mu\text{m}$  — the atmosphere would be too transparent. So methane clouds were suggested as an additional source of opacity. Analogies with terrestrial meteorology suggested that moist convection and cumulus clouds might be important in Titan's lower atmosphere.

But improved calculations gave gaseous opacities of nitrogen and methane that were higher than originally thought — high enough to explain the observed far-infrared spectra if the methane is somehow supersaturated<sup>13,14</sup> in the lower atmosphere, with a relative humidity of about 200%. This decidedly non-terrestrial situation could be accounted for by the lack of nucleation sites. On Earth, condensation nuclei abound, ranging from soluble salts from the oceans to airborne dust particles, whereas on Titan the inventory of available nuclei is probably much more limited, consisting perhaps of stratospheric aerosols making up the visible haze. Whether cumulus activity would occur in such an atmosphere is questionable. So direct evidence for clouds and their type is needed.

All of which brings us to the new work. On two nights in 1995, Griffith *et al.*<sup>1</sup> saw that Titan's disk was brighter than on previous occasions when the same central meridian faced Earth. The authors can explain this brightening as being due to a thick cloud at an altitude of roughly 15 km covering about 9% of Titan's disk. In later observations, in 1997, they saw no such brightening. This is the first direct indication of cloud activity on Titan. But the temporal resolution thus far is not very high, so one hopes for a systematic programme with more frequent observa-

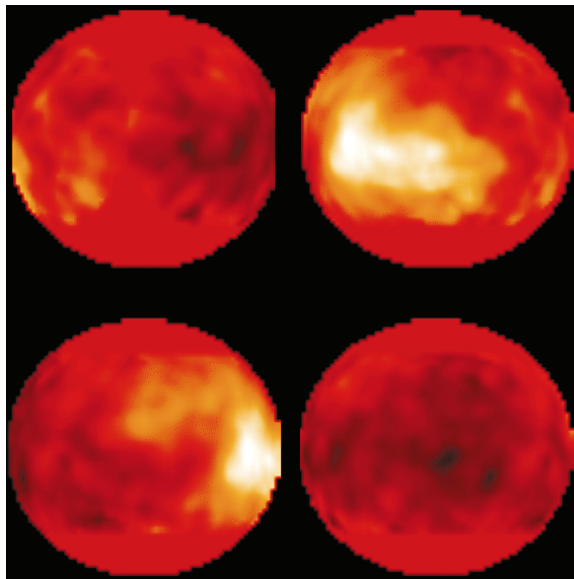


Figure 1 Four sides of Titan. These images show permanent surface structures, scuppering the idea of a global ocean of natural gas. But there is new evidence for methane clouds, perhaps part of a methane cycle similar to Earth's hydrological cycle.

NASA/HST

tions to confirm and better characterize the phenomenon.

The Cassini mission, a joint NASA/ESA effort, may fit the bill. Launched in the autumn of 1997, the Cassini spacecraft will arrive at Saturn in the summer of 2004. In the following November, the Huygens probe it carries will descend into Titan's atmosphere, armed with a camera to search for clouds as well as with a battery of instruments designed to measure temperatures, pressures, winds, gas composition (including methane), condensates and the properties of the surface.

For four years the spacecraft itself will orbit Saturn, making 40 or more close passes of Titan and employing instruments capable of observing from ultraviolet to radio wavelengths, including an onboard radar system that uses the high-gain antenna normally used for telecommunication and radio-occultation soundings. Data from the orbiter and the probe should clarify many issues about Titan, and will no doubt raise other perplexing questions. But the four-year life of the Cassini nominal mission is

short compared with Titan's year (29.5 Earth years) and several natural atmospheric timescales, so long-term coverage by ground-based and Earth-orbiting telescopes will still be important. □

*F. Michael Flasar is in the Laboratory for Extraterrestrial Physics, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. e-mail: mike@leprss.gsfc.nasa.gov*

1. Griffith, C. A., Owen, T., Miller, G. A. & Geballe, T. *Nature* **395**, 575–578 (1998).
2. Lindal, G. F. *et al. Icarus* **53**, 348–363 (1983).
3. Hunt, D. M. *et al. in Saturn* (eds Gehrels, T. & Matthews, M. S.) 671–759 (Univ. Arizona Press, 1984).
4. Hunt, D. M. *J. Atmos. Sci.* **30**, 726–732 (1973).
5. Strobel, D. F. *Planet. Space Sci.* **30**, 839–848 (1982).
6. Flasar, F. M. *Science* **221**, 55–57 (1983).
7. Lunine, J. I., Stevenson, D. J. & Yung, Y. L. *Science* **222**, 1229–1230 (1983).
8. Muhleman, D. O., Grossman, A. W., Butler, B. J. & Slade, M. A. *Science* **248**, 975–980 (1990).
9. Lemmon, M. T., Karkoschka, E. & Tomasko, M. *Icarus* **103**, 329–332 (1993).
10. Smith, P. H. *et al. Icarus* **119**, 336–349 (1996).
11. Sagan, C. & Dermott, S. F. *Nature* **300**, 731–733 (1982).
12. Samuelson, R. E., Hanel, R. A., Kunde, V. G. & Maguire, W. C. *Nature* **292**, 688–693 (1981).
13. Courtin, R., Gautier, D. & McKay, C. P. *Icarus* **114**, 114–162 (1995).
14. Samuelson, R. E., Nath, N. R. & Borysow, A. *Planet. Space Sci.* **45**, 959–980 (1997).

## Molecular endocrinology

# Two orphans find a home

Didier Picard

The superfamily of ligand-regulated transcription factors has grown to the point where the founding members — the steroid and thyroid hormone receptors — have become respected elders. To qualify for membership in this family, all that a protein needs is some homology to the DNA-binding or ligand-binding domains (or both) of a seasoned member. Being a transcription factor or, at least, being able to bind canonical DNA sequence elements is a

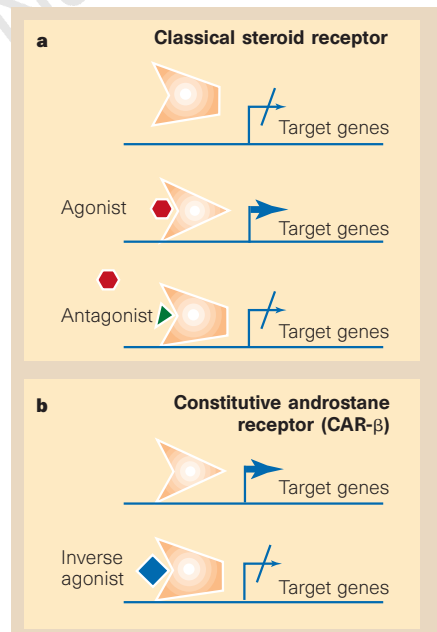
plus, but being a receptor for a known ligand is not a requirement. As a result, most members are considered 'orphan' nuclear receptors, waiting to be adopted by a ligand<sup>1</sup>. Finding parents, assuming there are any at all, has been tedious and rarely successful. But, on page 612 of this issue, Forman and colleagues<sup>2</sup> report such a success story with several unexpected twists.

When the constitutively active receptor (CAR) was cloned by Moore and colleagues

(a sub-set of the collaborating authors here), this orphan nuclear receptor was presumably given that name because of its constitutive transcriptional activity<sup>3</sup>. What incredible foresight in choosing this acronym — Forman *et al.*<sup>2</sup> now show that CAR is a constitutive androstane receptor (Fig. 1). Not only is it the receptor for two steroids, the androstane metabolites androstenol and androstanol, but, surprisingly, these steroids switch the activity of CAR off. Although ligand-deactivated repressors (such as the *lac* repressor) are known in prokaryotes, it is unusual to see this effect on a ligand-repressed transcription factor such as CAR. One example is the thyroid hormone receptor, which can be regulated similarly, but only on a rather unusual DNA response element<sup>4</sup>. It will be interesting to see how CAR's next of kin, the human orphan receptor MB67 (ref. 3; now called hCAR- $\alpha$  to distinguish it from the mouse CAR- $\beta$  discussed here), responds to androstane metabolites.

Does CAR- $\beta$  bind directly to androstenol and androstanol? Because radiolabelled ligands are often not available for ligand-binding assays, preliminary evidence is usually obtained indirectly. This is the case here. Transcriptionally active nuclear receptors associate with transcriptional co-activators, and Forman *et al.* found that, as expected, *in vitro* translated CAR- $\beta$  associates with the steroid-receptor co-activator-1 (SRC-1). Androstenol induces the dissociation of CAR- $\beta$  and SRC-1. This, and the fact that activity of CAR- $\beta$  is also repressed when assayed in yeast, suggests that androstenol and androstanol may not have to be further metabolized. But from the point of view of a receptor purist, this has yet to be shown with purified components. For example, perhaps these compounds are metabolized to the active form, even in yeast. Or maybe they signal through another component, present in both yeast and *in vitro* translation extracts. Moreover, CAR- $\beta$  has a dimerization partner, the retinoid-X receptor (RXR). So, does CAR- $\beta$  form the ligand-binding pocket by itself, or only in conjunction with RXR? RXR has been caught in very intimate liaisons, with the ecdysone receptor, for example, where both subunits are required to bind the ligand<sup>5</sup>.

Should the CAR- $\beta$  ligands be thought of as inverse agonists or as antagonists? This is not just semantics, because antagonists act against agonists, whereas inverse agonists simply have the opposite effect to an agonist. We cannot exclude the hypothesis of an endogenous ligand, although it would have to be present in mammalian cells, yeast and the *in vitro* system. Moreover, there does seem to be a precedent in the nuclear-receptor family. LXR- $\alpha$  is a nuclear receptor for oxysterols<sup>6</sup> that can be constitutively active, presumably owing to endogenous ligands. It can be inhibited by geranylgeraniol (but,



**Figure 1 Alternative pathways for signalling through intracellular receptors. a, Classical steroid receptor. The receptor cannot activate transcription unless it is activated by binding of an agonist. Antagonists can block this interaction and prevent transcription. b, The constitutive androstane receptor (CAR- $\beta$ ), described by Forman *et al.*<sup>2</sup>. The receptor is constitutively active, but binding of the androstane metabolites androstenol or androstanol turns the receptor off. (Note that sources for agonists, antagonists and inverse agonists can be autocrine or paracrine as well as endocrine.)**



#### 100 YEARS AGO

If it be permissible to judge from the books with which they are respectively supplied, there must be an inherent constitutional difference between the English and the American reader of popular bird-lore. In almost all the numerous works written for the benefit of the former there is a more or less rigid adherence to a systematic arrangement of some kind or other. As we have had previous occasion to remark, American books, on the other hand, are characterised by their partiality for methods of arrangement other than systematic. Personally we confess to a deep-rooted prejudice in favour of the English plan; but if American readers find this too cut and dried for them, and prefer some less inelastic classification, little fault can be found with the writers who endeavour to gratify their taste. From *Nature* 6 October 1898.

#### 50 YEARS AGO

A very simple experiment seems to have a meteorological significance. A small plate about one inch in area was attached to the insulated quadrant of a Dolezalek electrometer. A cupful of fairly dry sand was then poured through a funnel and allowed to fall some three feet to the floor. The electrometer plate was three yards away from the dropping sand. The passage of the sand through the funnel took about half a minute, and while it was dropping there was no effect on the electrometer; but soon after, the electrometer began to show a deflexion which continued to increase for three or four minutes. This deflexion corresponded to the plate receiving positive charge. ... The meteorological application is, of course, obvious. When sand is blown about on a big scale, very large charges must be produced, and the different rates of diffusion or convection separate the charges. ... Many years ago, on the Salonika front, I noticed that when a dust storm was approaching my wireless station the aërials started sparking violently when the dust cloud was some way off; sparking then ceased for a while, to recommence with equal violence when the storm hit the camp. This has an uncommon resemblance to the electrometer experiment, the first sparking being due to particles too fine to be visible. From *Nature* 9 October 1948.

again, is this an inverse agonist or an antagonist?'). What we know about the structure-function relationship of ligand-binding domains should help us to design CAR- $\beta$  mutants that cannot bind a hypothetical endogenous ligand, and are resistant to the inverse agonists.

One would hope that structural biologists will quickly adopt CAR- $\beta$ . The carboxy-terminal helix, which contributes to the interaction surface with transcriptional co-activators in other nuclear receptors, is probably in the 'on' position when CAR- $\beta$  is not bound to the ligand. Perhaps, on binding ligand, this helix is displaced — in the same way as binding of the anti-oestrogen raloxifen to the oestrogen receptor<sup>8</sup> causes displacement. The structural analysis might also help us to guess which orphan nuclear receptors are the most promising candidates in a hunt for inverse agonists.

What about the ligands? So little is known about their biology or biochemistry that they could hardly have been suspected to be orphan ligands at all. Although Forman

*et al.*<sup>2</sup> find that the stereospecificity of androstanol and androstenol is remarkable — all related compounds were inactive — we cannot yet be sure that they really are the biological parents of the orphan CAR- $\beta$ . As I have previously discussed in these columns<sup>9</sup>, Forman *et al.* confirm that many steroid metabolites are not just precursors at best, and cellular rubbish at worst, but that they can be signals in their own right. □

Didier Picard is in the Département de Biologie Cellulaire, Université de Genève, Sciences 3, 1211 Genève 4, Switzerland.  
e-mail: Picard@cellbio.unige.ch

1. Enmark, E. & Gustafsson, J.-Å. *Mol. Endocrinol.* **10**, 1293–1307 (1996).
2. Forman, B. M. *et al. Nature* **395**, 612–615 (1998).
3. Choi, H. S. *et al. J. Biol. Chem.* **272**, 23565–23571 (1997).
4. Saatcioglu, F., Deng, T. & Karin, M. *Cell* **75**, 1095–1105 (1993).
5. Yao, T. P. *et al. Nature* **366**, 476–479 (1993).
6. Janowski, B. A., Willy, P. J., Devi, T. R., Falck, J. R. & Mangelsdorf, D. J. *Nature* **383**, 728–731 (1996).
7. Forman, B. M., Ruan, B., Chen, J., Schroepfer, G. J. Jr & Evans, R. M. *Proc. Natl Acad. Sci. USA* **94**, 10588–10593 (1997).
8. Brzozowski, A. M. *et al. Nature* **389**, 753–758 (1997).
9. Picard, D. *Nature* **392**, 437–438 (1998).

#### Photochemistry

## Solid-state organic solar cells

Wim C. Sinke and Martijn M. Wienk

A promising new type of solar cell is described by Bach *et al.* on page 583 of this issue<sup>1</sup>. More efficient than other solid-state organic devices, it could turn out to be a prototype for cheap, large-area cells.

Ever since the invention of the silicon solar cell in the 1940s, people have acknowledged the enormous potential that photovoltaic systems have for large-scale electricity production. But semiconductor-grade silicon wafers are expensive, so great efforts have been put into the development of cheaper thin-film solar cells and modules. Such films may be purely inorganic (amorphous silicon,

cadmium telluride, copper-indium-diselenide), or contain organic materials as an essential part of the device. Examples of the latter are junctions consisting of thin layers of organic donor (D) and acceptor (A) dye molecules<sup>2</sup>; bulk junctions in which the donor and acceptor phases exist as an interpenetrating network<sup>3</sup>; and dye-sensitized photoelectrochemical solar cells, in which light absorption occurs in metallo-organic dye molecules, but the generated charge carriers are transported by an inorganic semiconductor such as TiO<sub>2</sub>.

The interest in dye-sensitized photoelectrochemical solar cells has increased enormously since the 1991 report<sup>4</sup> on the application of 'nanostructured' TiO<sub>2</sub>, which has pores on a scale of nanometres. Because of the large surface area of this material, a monolayer of dye (which ensures very efficient charge separation between dye and TiO<sub>2</sub>) absorbs enough light to achieve useful efficiencies. In fact, these devices have shown record efficiencies of 11% at 0.25 cm<sup>2</sup> and 6.5% at 1.6 cm<sup>2</sup> (ref. 5).

But although the initial efficiency of small individual cells is of great scientific importance, and high cell efficiencies are a stimulus for technological development, most commercial applications require efficient, large-area multi-cell modules that are simple to make and stable in the long term. So several companies and institutes have concentrated their efforts on these aspects of dye-sensitized photovoltaic devices.

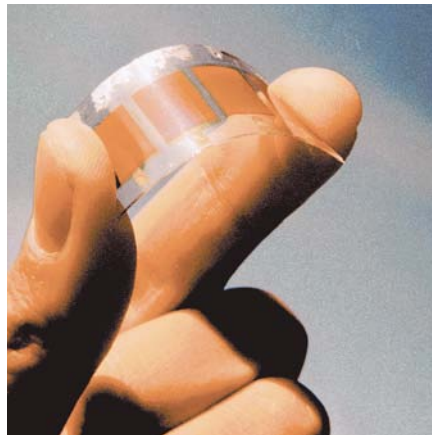


Figure 1 Prototype of a flexible organic solar cell that uses a liquid electrolyte. An amorphous conducting material has now been used as an alternative to the electrolyte in an otherwise comparable device.



The use of a liquid electrolyte, as in the original cell design<sup>4</sup>, may limit device stability because the liquid may evaporate when the cell is imperfectly sealed, and more generally diffusion and reaction of water or oxygen molecules may worsen cell performance. The liquid electrolyte also makes the manufacture of multi-cell modules difficult because cells must be connected electrically yet separated chemically, preferably on a single substrate. One way to remove some of these problems is to replace the liquid electrolyte with a solid conducting material. Researchers have tried polymer electrolytes that conduct ions<sup>6</sup> or inorganic materials such as CuI and CuSCN that conduct holes<sup>7,8</sup>. But so far, the devices produced have performed poorly.

The new approach of Bach *et al.*<sup>1</sup> is to use an amorphous conducting material. They have developed an organic hole-transport material (HTM), in which positive charges hop from one molecule to another. The material can be applied simply, by spin casting, and because the two halves of the molecule are perpendicular to each other (Fig. 1 on page 583), a three-dimensional structure is imposed that results in an amorphous solid state, without the grain boundaries that limit hole transport. Furthermore, the amorphous state is expected to form a better interface with the dye-covered TiO<sub>2</sub> electrode than would a crystalline state, ensuring efficient hole transport from the oxidized dye to the HTM. This has resulted in cells where up to 30% of incident photons are converted into a charge carrier. Although this value is still appreciably lower than the maximum conversion fraction of liquid electrolyte cells (up to 90%), it is very high for an organic solid device.

The open-circuit voltage of 342 mV, however, is much lower than the value that might be expected. That may indicate recombination losses, possibly owing to the partial doping of the HTM needed to obtain high enough conductivity. To achieve high efficiency, solid-state, dye-sensitized solar cells, this is probably the biggest obstacle to be overcome. Furthermore, it remains to be seen whether the sensitivity of such solid-state devices to water or molecular oxygen is lower than that of liquid-based devices, and whether the ultraviolet stability is good enough.

Nevertheless, Bach *et al.* have taken the development of solid-state, dye-sensitized solar cells a stage further. These cells are still in their early infancy, particularly compared with inorganic thin-film devices, but the promise of very low production costs and reasonable to high efficiencies warrants intensive work in this area, even if it is many years before the resulting solar modules cover our roofs. □

Wim C. Sinke and Martijn M. Wienk are at the Netherlands Energy Research Foundation (ECN), POB 1, NL-1755 ZG Petten, The Netherlands.

e-mails: sinke@ecn.nl

wienk@ecn.nl

1. Bach, U. *et al.* *Nature* **395**, 583–585 (1998).
2. Wöhrlé, D. *et al.* *J. Mater. Chem.* **5**, 1818–1829 (1995).
3. Yu, G., Gao, J., Hummelen, J. C., Wudl, F. & Heeger, A. J. *Science* **270**, 1789–1791 (1995).
4. O'Regan, B. & Grätzel, M. *Nature* **353**, 737–739 (1991).
5. Green, M. A., Emery, K., Bücher, K., King, D. L. & Igari, S. *Prog. Photovoltaics: Res. Appl.* **6**, 35–42 (1998).
6. Cao, F., Oskam, G. & Searson, P. C. *J. Phys. Chem.* **99**, 17071–17073 (1995).
7. Tennakone, K., Kumara, G. R. R. A., Kumarasinghe, A. R., Wijayantha, K. G. U. & Sirimanne, P. M. *Semicond. Sci. Technol.* **10**, 1689–1693 (1995).
8. O'Regan, B. & Schwartz, D. T. *Chem. Mater.* **10**, 1501–1509 (1998).

## Floral patterning

# A LEAFY link from outer space

Ben Scheres

The beautiful shape and colour patterns of flowers attract almost everybody, including developmental biologists. Studies of floral patterning have shown that the identity of floral organs is determined by a combinatorial code of homeotic genes, termed A, B and C, expressed in three separate regions of the flower<sup>1</sup>. This knowledge allows scientists (and biotech companies) to manipulate flowers to form organs at any position, yet the secret of how the three activities are put into the right place remains hidden. But on page 561 of this issue, Parcy *et al.*<sup>2</sup> bring new insights to this matter. They show that the LEAFY (LFY) protein is directly involved in activating the transcription of homeotic genes, together with at least one co-regulator, UNUSUAL FLORAL ORGANS (UFO). The genes that encode these proteins were originally isolated as

*floricaula* and *fimbriata* in a flower that almost all of us have manipulated as children, the snapdragon<sup>3,4</sup>.

Parcy and colleagues used *Arabidopsis thaliana* for their study. *Arabidopsis* flowers are formed on the flanks of inflorescence meristems — groups of stem cells that give rise to flower-carrying branches. Inflorescence meristems, in turn, derive from the shoot apical meristem. This is first laid down in the plant embryo, and it produces leaves in the vegetative phase before switching to the floral programme (Fig. 1). Within floral meristems, the sepals, petals, stamens and carpels are specified in concentric rings: A genes specify sepals; A and B genes, petals; B and C genes, stamens; and C genes alone, carpels. In *Arabidopsis*, A-gene function is carried out by the *APETALA1* (*API*) and *AP2* genes, B-gene function by *AP3* and *PISTIL-*

*LATA*, and C-gene function by *AGAMOUS* (*AG*).

The *LFY* gene is a meristem-identity gene — it promotes the formation of floral meristems<sup>3,5</sup>. The gene product, LFY, is a DNA-binding nuclear protein that is expressed throughout the floral meristem<sup>2</sup>. Parcy *et al.* now show that LFY is also able to activate homeotic genes, and that this activity can be uncoupled from its meristem-identity function. These findings were primed by analysis of *lfy* mutants. Many *lfy* mutant flowers are transformed into inflorescence meristems, but some flowers eventually form, with abnormalities that suggest differential effects of LFY on the activity of homeotic genes. Parcy *et al.*<sup>2</sup> reasoned that the ability of LFY to activate homeotic genes

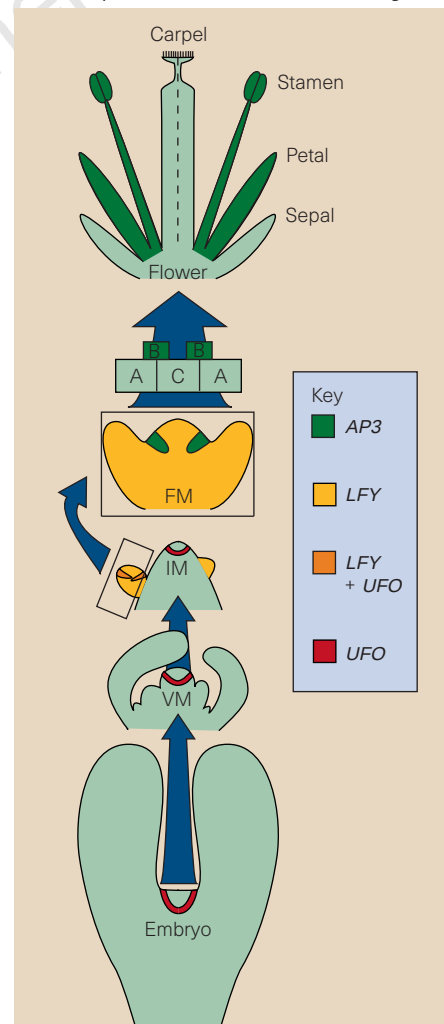


Figure 1 Parcy and colleagues' model<sup>2</sup> for B-region patterning by a combination of spatial information from UNUSUAL FLORAL ORGANS (UFO) and flower-meristem specificity from LEAFY (LFY). There is patterned distribution of UFO RNA (red) from embryogenesis onward, then LFY protein (yellow) is expressed throughout stage-1–3 floral meristems, and the B-function gene *APETALA3* (*AP3*; green) is expressed in the region of overlap. VM, vegetative (shoot apical) meristem; IM, inflorescence meristem; FM, floral meristem.

might be regulated differentially, to allow region-specific expression of these genes. Indeed, they concluded that LFY alone can initiate expression of *API* in the A region; that it requires an unknown cofactor to activate expression of *AG* in the C region; and that *UFO* acts together with LFY to activate *AP3* in the B region (Fig. 1).

To investigate whether LFY is directly involved in floral patterning, Parcy *et al.* made a constitutively active LFY variant by fusing a transcriptional-activator domain to the LFY gene. Transgenic plants carrying this construct formed floral meristems at the normal times and places, indicating that the early function of LFY was not altered. However, these transgenic plants expressed *AG* outside the flowers (ectopically), and early expression of *API* was increased. Whereas ectopic transcription of *API* could be induced by expression of LFY alone, induction of ectopic *AG* required expression of the activated form of LFY. This indicates that induction of *AG*—but not *API*—requires a flower-specific cofactor. Moreover, although neither LFY nor activated LFY could induce ectopic transcription of *AP3*, constitutive expression of the *UFO* gene with LFY could induce ubiquitous transcription of *AP3*. Promoter-binding and protein-interaction studies should now corroborate, at the molecular level, the regulatory interactions inferred from the LFY gene fusion.

Activation of *AG* leads to repression of *API*, thereby restricting *API* to the A region (that is, the petals and sepals)<sup>6</sup>. Because activation of *AG* in the C region leads to patterning of the A region, the postulated LFY co-regulators in the B and C regions form the minimal requirement to pattern flowers. As well as being involved in spatial control (along with the proposed co-regulators), LFY integrates temporal input from 'flowering time' genes, such as *CONSTANS*, which convey environmental information<sup>7</sup>.

Additional thrilling news reported by Long and Barton<sup>8</sup> in *Development* is that *UFO* is expressed not only in a B-region pattern in floral meristems, but also in shoot meristems from embryogenesis onwards (Fig. 1). Furthermore, genes with shoot-meristem functions, such as *CLAVATA1*, are already expressed in a C-region-like pattern in the embryo<sup>9</sup>. These genes may provide, or respond to, cues from the C region. Thus, flowers can read out position from cues that are established in every shoot-derived meristem from embryogenesis onwards. In evolutionary terms, the floral pattern seems to hitchhike on spatial information that was in place in the more ancient shoot apical meristem.

Sceptics might raise their eyebrows at this point—we already know from animal development that new patterns can arise from pre-existing ones. However, the iterative initiation of patterning cues, which can serve different

purposes under the control of environmental inputs, is a new situation. Moreover, a conceptual framework for plant patterning has not emerged in a single sweep, as it has in the fruitfly<sup>10</sup>. In plants, no fly equivalent is available, and genetic redundancy promises to be as omnipresent as in vertebrates. The repetitive nature of development further complicates analysis. All this makes one increasingly aware of the significance of this new link between flower and embryo. □

Ben Scheres is in the Department of Molecular Cell Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.  
e-mail: b.scheres@bio.uu.nl

1. Bowman, J. L., Smyth, D. R. & Meyerowitz, E. M. *Development* **112**, 1–20 (1991).
2. Parcy, F., Nilsson, O., Busch, M. A., Lee, I. & Weigel, D. *Nature* **395**, 561–566 (1998).
3. Coen, E. S. *et al.* *Cell* **63**, 1311–1322 (1990).
4. Simon, R. *et al.* *Cell* **78**, 99–107 (1994).
5. Weigel, D. *et al.* *Cell* **69**, 843–859 (1992).
6. Gustafson-Brown, C., Savidge, B. & Yanofsky, M. F. *Cell* **76**, 131–143 (1994).
7. Simon, R., Igerio, M. I. & Coupland, G. *Nature* **384**, 59–62 (1996).
8. Long, J. A. & Barton, M. K. *Development* **125**, 3027–3035 (1998).
9. Clark, S. E., Williams, R. W. & Meyerowitz, E. M. *Cell* **89**, 575–585 (1997).
10. Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795–801 (1980).

## Animal evolution

# From deep time to late arrivals

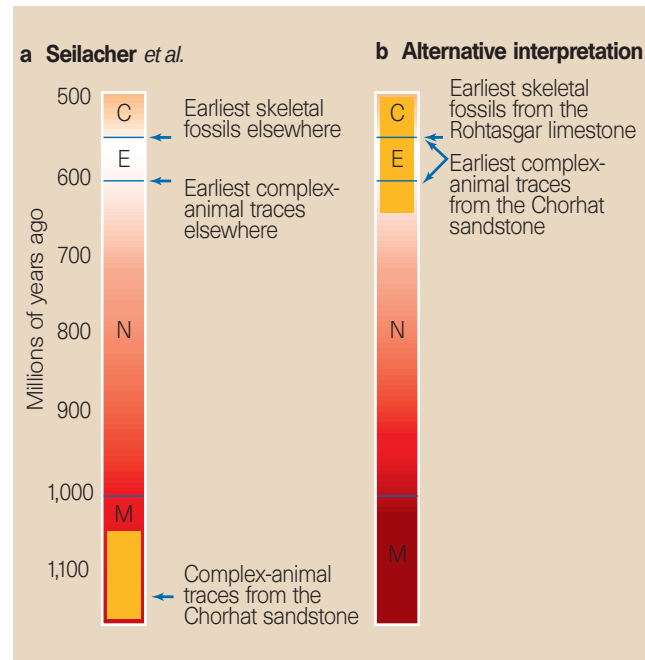
Martin Brasier

In a paper published last week in *Science*<sup>1</sup>, Adolf Seilacher and colleagues claim that they have found evidence of complex, multicellular animals in sediments generally thought to be more than 1,100 million years old. The sediments concerned are the Chhorhat sandstones of the Vindhyan super-group in central India. The authors report markings from the tops of bedding planes which, they suggest, are tunnels made by small worms burrowing just beneath the surface of bacterial mats. If these markings are accepted as evidence for the existence of animals more than a billion years ago, they would double the time span of the animal fossil record.

One context in which the report should be seen is what was thought to be a mismatch between the evidence provided by fossils and that provided by 'molecular clocks'. Hitherto, the oldest traces of complex-animal

activity have been in the region of 550–600 million years old<sup>2,3</sup>. According to some researchers, however, this is much younger than evidence from analysis of the ribosomal RNA of living organisms would suggest. Assuming that gene sequences evolve with such regularity that differences can be used as 'molecular clocks', it can be argued that invertebrate lineages began to diverge about 1,200 million years ago<sup>4</sup>. This would mean that the earliest animals are missing from the fossil record because they left few, if any, remains. For example, they may have lived in the water column or as microscopic life in sediments. The so-called Cambrian 'explosion' 600 million years ago would then largely relate to the acquisition of skeletons.

Molecular clocks are notoriously difficult to calibrate, however, and these 'deep time' estimates for the divergence of the major groups of animals have been drastically



**Figure 1 Time and animal evolution.** a, Seilacher and colleagues' view. Based on the authors' interpretation<sup>1</sup> of possible trace fossils in the Chhorhat sandstone, the earliest evidence of complex animals appears about 1,100 million years ago, in the Mesoproterozoic (M) and well before the Neoproterozoic (N), Ediacaran (E) and Cambrian (C). b, An alternative interpretation, in which the trace fossils are only slightly older than newly discovered Cambrian skeletal fossils from the overlying Rohtasgar limestone<sup>10</sup>. The age span of the Vindhyan rocks is shown in yellow.

scaled down to nearer 670 million years<sup>5</sup>, bringing the figures more into line with the fossil record. A 'late arrival' model would also imply that the evolution of the main animal groups took place both late in Earth's history, and rapidly — perhaps in response to the evolution of Hox genes<sup>6</sup> or to the lifting of some external constraint such as insufficient atmospheric oxygen<sup>2</sup>.

So should these supposedly ancient markings be accepted as new evidence for the 'deep time' model? When a respected researcher such as Seilacher, who has a good track record in the field of trace fossils, makes such a claim as this, it has to be taken seriously. But extraordinary claims, of course, require extraordinary evidence.

The first claim<sup>1</sup> is that these are the traces of 'triploblastic' animals. This means that the trace-makers had a gut and a fluid-filled coelom. Such an organization is found only in complex animals from worms to chordates. Seilacher and co-authors' argument is that, at 5 mm, the burrow diameters are too large to have been made by single-celled protists; but unicellular protists, including early Cambrian forms, can be that large<sup>7</sup> and can make burrows. Moreover, at 5 mm diameter, how could such organisms have remained hidden from the fossil record for over 500 million years? Finally, how come the tunnels are branched? Branching implies a behavioural sophistication which is thought to have appeared about 560 million years ago<sup>8</sup>.

These seeming paradoxes may yet be explained by a fresh look at the age of the lower Vindhyan rocks (Fig. 1). Until now, they have broadly been taken to be 1,100 million years old on the basis of potassium-argon and fission-track dates. But these dates were mostly obtained in the 1960s, and no further particulars are given by the authors<sup>1</sup>. Unfortunately, the data they cite must now be regarded as a very doubtful guide to the age of the Vindhyan rocks<sup>9,10</sup>, for which modern studies of geochronology are urgently required.

In fact, we may be in for a big surprise regarding the age of the Vindhyan supergroup, including, of course, the Chorhat sandstone. Acid digestion of the immediately overlying Rohtasgar limestones, undertaken in India by R. J. Azmi, has yielded a rich, well-preserved and typical early Cambrian skeletal fauna, including certain types of brachiopod<sup>10</sup>. This astounding discovery means that supposed trace-fossil markings from the Chorhat sandstone may be little more than 540 million years old, close to the beginning of the Cambrian 'explosion'. Branched burrow systems up to 5 mm wide are certainly well known from the Ediacaran and Cambrian<sup>8</sup>. Unless there is a major break between the Chorhat sandstone and the Rohtasgar limestone, this new interpretation<sup>10</sup> would mean that there is

still no reliable fossil evidence to support the 'deep time' view of the emergence of animals. □

Martin Brasier is in the Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.

e-mail: martin.brasier@earth.ox.ac.uk

1. Seilacher, A., Bose, P. K. & Pflüger, F. *Science* **282**, 80–83 (1998).
2. Schopf, J. W. & Klein, C. (eds) *The Proterozoic Biosphere* (Cambridge Univ. Press, 1992).
3. Brasier, M. D. & McLroy, D. J. *J. Geol. Soc. Lond.* **155**, 5–12 (1998).

4. Wray, G. A., Lewinton, J. S. & Shapiro, L. H. *Science* **274**, 568–573 (1996).
5. Ayala, F. J. & Rzhetsky, A. *Proc. Natl Acad. Sci. USA* **95**, 606–611 (1998).
6. Erwin, D., Valentine, J. & Jablonski, D. *Am. Sci.* **85**, 126–137 (1997).
7. McLroy, D., Green, O. R. & Brasier, M. D. *Microscopy and Analysis* 113–115 (November 1994).
8. Crimes, T. P. in *The Palaeobiology of Trace Fossils* (ed. Donovan, S. K.) 105–133 (Wiley, New York, 1994).
9. Venkatachala, B. S., Sharma, M. & Shukla, M. *Mem. Geol. Soc. Ind.* **36**, 137–165 (1996).
10. Azmi, R. J. *J. Geol. Soc. Ind.* (in the press).

## RNA structure

# Ribozyme crevices and catalysis

Daniel Herschlag

Since the X-ray crystallographic structure of the hammerhead ribozyme — a catalytic RNA molecule — was solved<sup>1</sup> in 1994, the RNA community has been waiting for a view of a second RNA enzyme. We are now rewarded for our patience by not one but two ribozyme structures: a 2.3-Å structure of the ribozyme from the hepatitis delta helper virus (HDV), reported by Ferré-D'Amaré *et al.*<sup>2</sup> on page 567 of this issue; and a 5-Å structure of the group I intron from *Tetrahymena thermophila*, published by Golden *et al.*<sup>3</sup> in this week's *Science*. The most striking feature of these RNAs is that they look like enzymes — that is, both have active sites that seem to be preorganized and located in crevices, as we are accustomed to seeing in protein enzymes. This is in marked contrast to the hammerhead ribozyme, which exposes the cleavage site in solution, rather than sequestering it in a cavity, and seems to require a substantial conformational rearrangement to perform its catalytic function<sup>4,5</sup>.

Despite their general similarities, the HDV and *Tetrahymena* ribozymes are preorganized very differently. The HDV ribozyme is only about one-fifth the size of the *Tetrahymena* ribozyme, owing to a remarkable intertwining of secondary structural elements in what can be referred to as a 'double pseudoknot'. The positioning of secondary structure (that is, base-pairing) — which accounts for three-quarters of the HDV residues — within the double pseudoknot establishes the overall position of the helices with respect to each other. This allows a complex, three-dimensional architecture. Several tertiary connections then establish the precise tertiary fold, positioning the functional groups that are required for catalysis, despite the ribozyme's modest size.

In contrast, group I ribozymes are constructed from domains, some of which constitute independent folding units. The structure of one of these domains, the P4–P6 domain, has already been solved to high resolution<sup>6</sup>, and this domain is largely unper-

turbed in Golden and colleagues' structure. The second domain, P3–P9, was not present in the previous structure, but it can now be seen to wrap around the P4–P6 domain. This observation provides a structural explanation for kinetic-folding results<sup>7,8</sup>, which show that the P4–P6 domain forms earlier than the P3–P9 domain.

This structure, along with the many functional data for the *Tetrahymena* ribozyme, shows how the coming together of these domains is critical for forming the substrate binding sites. The binding site for one substrate, guanosine, was previously localized biochemically<sup>9</sup> to the P7 helix, which is found within the P3–P9 domain. The structure reveals that the P7 helix is distorted when P3–P9 interacts with the P4–P6 interface, presumably allowing P7 to bind guanosine. Distortion of the P7 helix explains an observation from my own laboratory (M. Engelhardt and D. H., unpublished) that removal of the P5abc subdomain of P4–P6 weakens guanosine binding, even though P5abc and P7 are not in direct contact. The intersection of the P4–P6 and P3–P9 domains also creates a shallow crevice. This is lined with residues that have been implicated in binding of the substrate domain P1–P2, which is absent from the molecule crystallized by Golden *et al.*<sup>3</sup>. Further experiments will reveal whether the 5-Å limit to resolution comes from limits that are specific to this crystal, or whether it is due to inherent flexibility in the absence of the P1–P2 substrate domain.

The two new structures provide an opportunity to evaluate what works in structural prediction, what doesn't work, and why. Because RNA secondary structure is determined locally, most secondary-structure elements can be identified through evolutionary relationships and mutagenic tests involving compensatory changes of putative pairing partners. A remarkably accurate three-dimensional model of the group I intron was derived from such analysis<sup>10,11</sup>. Similarly, secondary-structure prediction turns out to have been very good for



HDV. In addition, the global placement of helices was correctly predicted from the constraints on the secondary structure, and groups identified as critical in functional assays are now seen to make central structural connections or inhabit the active-site crevice<sup>12</sup>. However, helix P1.1, which was revealed in Ferré-D'Amaré and colleagues' structure, was not identified previously. Moreover, a three-dimensional structure obtained by modelling does not fully reflect the observed tertiary structure.

One practical approach towards understanding the value and limits of a model would be to adopt, in three dimensions, what RNA modellers are already doing in two — deriving a family of possible low-energy structures. Regions of similarity in the structures would suggest interactions with strong constraints, whereas regions with substantial differences might highlight important features that require additional constraints. Biochemical tests might then distinguish between the possibilities.

Moderate- and high-resolution structures allow us to identify motifs and the overall architecture of an RNA molecule, and can also provide a powerful framework for interpreting biochemical results. In addition, the atomic-level pictures from high-resolution structures, such as that of the HDV ribozyme and the P4–P6 domain of the *Tetrahymena* ribozyme<sup>2,6</sup>, mean that we can use biochemical experiments to probe the energetics of individual interactions and address specific questions about structure and folding.

But what about catalytic mechanism? What are the interactions of the ribozyme with the transition state, the highest-energy structure along the reaction pathway that must be stabilized by a catalyst? The active sites for protein enzymes are typically found in crevices. This allows the protein to position functional groups precisely, through networks of interactions, and to interact with substrates. Although different in molecular detail, RNA enzymes seem to follow the same strategies<sup>13</sup>. The preorganized functional groups within the crevice can provide catalysis by positioning substrates and catalytic groups with respect to one another, and by electrostatic interactions that are strengthened in the reaction's transition state. These groups may also facilitate bond rearrangements, acting, for example, as general acid and base catalysts.

The functional groups that line the HDV and *Tetrahymena* ribozyme crevices are implicated as candidates for catalysis. But the nature of their transition-state interactions and catalytic roles is not clear, even for HDV where the structural certainty is greatest. In the HDV-catalysed reaction, do metal ions interact with the nucleophilic 2'-oxygen atom, with the departing 5'-oxygen atom or with the non-bridging oxygens of the transferred phosphoryl group? Are there hydro-

gen-bonding interactions with ribozyme functional groups or associated water molecules that position the reacting groups and stabilize charge redistribution in the transition state? In this regard, even for protein ribonucleases, where many structures to resolution beyond 2 Å are known (including structures with bound transition-state analogues), we still do not completely understand the catalytic mechanisms and contributions. Indeed, there is still discussion about the pathway of proton transfers within these reactions. In this light, it would be far too much to expect a single atomic-resolution structure to solve a mechanism. Nevertheless, the HDV structure narrows the list of suspects for direct catalytic interactions, and provides unrivalled power to suggest potential mechanisms. Future structural studies that locate the reactive phosphoryl group and the metal ions required for efficient catalysis, and probe the extent to which functional groups are rearranged within the active-site crevice, will help to direct mechanistic proposals and tests of these proposals.

The mysteries of function are brought into much sharper focus through the atomic

detail that structure provides. Mechanistic investigations and structural studies will accompany one another well into the next millennium, on the road to unravelling these mysteries. □

Daniel Herschlag is in the Department of Biochemistry, Beckman Center, Stanford University, Stanford, California 94305-5307, USA.  
e-mail: herschla@cmgm.stanford.edu

1. Pley, H. W., Flaherty, K. M. & McKay, D. B. *Nature* **372**, 68–74 (1994).
2. Ferré-D'Amaré, A. R., Zhou, K. & Doudna, J. A. *Nature* **395**, 567–574 (1998).
3. Golden, B. L., Gooding, A. R., Podell, E. R. & Cech, T. R. *Science* **282**, 259–274 (1998).
4. McKay, D. B. *RNA* **2**, 395–403 (1996).
5. Peracchi, A., Karpeisky, A., Maloney, L., Beigelman, L. & Herschlag, D. *Biochemistry* (in the press).
6. Cate, J. H. *et al.* *Science* **273**, 1696–1699 (1996).
7. Zarrinkar, P. P. & Williamson, J. R. *Science* **265**, 918–924 (1994).
8. Slavi, B., Sullivan, M., Chance, M. R., Brenowitz, M. & Woodson, S. A. *Science* **279**, 1940–1942 (1998).
9. Michel, F., Hanna, M., Green, R., Bartel, D. & Szostak, J. W. *Nature* **342**, 391–395 (1989).
10. Michel, F. & Westhof, E. *J. Mol. Biol.* **216**, 585–610 (1990).
11. Lehnert, V., Jaeger, L., Michel, F. & Westhof, E. *Chem. Biol.* **3**, 993–1009 (1996).
12. Been, M. D. & Wickham, G. S. *Eur. J. Biochem.* **247**, 741–753 (1997).
13. Narlikar, G. J. & Herschlag, D. *Annu. Rev. Biochem.* **66**, 19–59 (1997).

#### Lunar exploration

## Water amongst the rock

Timothy D. Swindle

The Moon isn't wet, but there is at least some water there. The latest and best evidence for water ice at the poles of the Moon comes from Feldman *et al.*<sup>1</sup> in a paper in *Science*. These findings may or may not make lunar bases more feasible, but they do represent a success for a new type of exploration for the US space agency NASA.

Although early astronomers called the vast, dark, circular patches of the Moon 'maria', thinking that they might be seas, water is scarce on the Moon. The amount of water in the samples returned by the Apollo astronauts is as low as 1 part per million, and most of that is probably contamination introduced during collection and processing<sup>2</sup>. Some water should be brought to the Moon in asteroids and comets, and some should be created near the surface by the reduction of iron-bearing minerals by hydrogen in the solar wind. But these processes would put the water near to the lunar surface. With daytime temperatures approaching 400 K in the regions spacecraft have visited, that water has little chance of staying there.

There are, however, areas of the Moon that are permanently shadowed from sunlight. These regions of perpetual darkness, primarily the interiors of impact craters at very high latitudes, can stay very cold (<120 K), so water might be retained there<sup>3,4</sup>. Such

regions should also act as cold traps, and collect water that is released from other places on the Moon. The idea that the lunar poles might harbour ice gained more attention with radar studies in the early 1990s that presented evidence for the presence of ice in polar regions of Mercury<sup>5–7</sup>, the closest planet to the Sun, which experiences even higher equatorial daytime temperatures than the Moon. But the permanently shadowed regions of the Moon have never been visited by spacecraft, and are difficult to observe from Earth because the same crater walls that block the sunlight often block the view from Earth.

Less than two years ago, the Clementine mission made headlines when data it had collected showed indications of ice at the lunar poles<sup>8</sup>. As in the case of Mercury, the evidence came from radar studies. Instead of looking at the radar return from a location, the Clementine studies consisted of analysing the scattering of the spacecraft's radio signal to Earth when it was bounced off the region of interest. Although they were exciting, the Clementine results were controversial, and later ground-based radar studies failed to confirm the presence of ice<sup>9</sup>.

Into the breach came the Lunar Prospector. It was built for only \$63 million, one to two orders of magnitude less than the planetary exploration spacecraft of the 1970s and

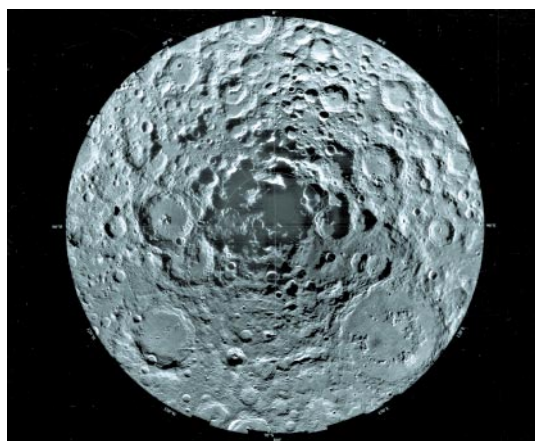


Figure 1 Dark and cold — and probably icy (at least beneath the surface). This is a mosaic of images of the Moon's south pole, taken by the Clementine spacecraft. Clementine's radar detection of water ice at the lunar poles was controversial. The data on neutron fluxes, collected by Lunar Prospector and discussed here, are much more compelling.

1980s, and was launched only 22 months after the decision to fly it was made<sup>10</sup>. Lunar Prospector is the first competitively selected mission (and the third overall) in NASA's Discovery Program, which emphasizes smaller, cheaper spacecraft with narrowly focused objectives. One of the reasons Prospector succeeded was because of a commitment to fly instruments that could meet specific goals (primary among them being the search for water at the lunar poles), rather than cutting-edge technology.

Lunar Prospector used a more direct approach to searching for ice than Clementine, and has achieved less ambiguous results. The detection was based on the measurement of the abundance and energy spectrum of neutrons coming from the interaction of galactic cosmic rays with the material in the uppermost metre of the lunar surface. As the lightest of the elements, hydrogen is uniquely capable of reducing the number of epithermal (moderate energy) neutrons. Feldman *et al.*<sup>1</sup> found a distinct lack of epithermal neutrons at each pole, strong evidence for the presence of hydrogen. Hydrogen will also reduce the number of neutrons with higher energies, but so will heavier elements, and these can mask the effects of hydrogen, if the hydrogen is buried. Because there was not a dramatic difference in the number of high-energy neutrons at the poles, Feldman *et al.* concluded that the hydrogen is buried, probably by 40 cm or more. A detection of hydrogen is not a detection of water; it could also be as H<sub>2</sub> or some other compound. But H<sub>2</sub>O is by far the most likely form, in part because the distribution matches that predicted for water. Ironically, although Lunar Prospector did apparently find ice at the lunar poles, it really did not confirm the earlier Clementine reports. Radar can detect ice only very near the surface. The neutron measurements suggest that the ice detected in that way is buried, so Feldman *et al.* say that their results are "consistent with" the null results of ground-based radar.

Estimating the total amount of water present at the lunar poles is not so easy. Feldman *et al.* give a best estimate of about six

billion tonnes. But several assumptions go into the calculation, including the purity of the ice and its vertical and spatial extent, and the figure could be an order of magnitude higher or lower. One way to improve the estimate is to acquire more data with higher spatial resolution, and Lunar Prospector will do this when its orbital height is reduced from 100 km to 40 km in December. Another way to improve the estimates is to determine the depth distribution more precisely. Although more Prospector data will help, it will probably require measurements on the lunar surface, or of samples returned to Earth, to truly understand the distribution.

The implications of all this go beyond science. An external source of water would be one of the most valuable resources for self-sustaining space exploration and colonization<sup>11</sup>. Moreover, the hydrogen is prized for rocket fuel, and the oxygen for breathing. If there are vast deposits of water at the lunar poles, it might be more economical to build a lunar base there than at the equatorial locations where spacecraft landings have been to date. On the other hand, the difficulties of working at such low temperatures might mean that the lunar poles would still be far less economical than other locations, such as near-Earth asteroids<sup>12</sup>. But at the least, the Lunar Prospector results keep the debate open. □

Timothy D. Swindle is in the Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721-0092, USA.

e-mail: [tswindle@U.Arizona.edu](mailto:tswindle@U.Arizona.edu)

1. Feldman, W. C. *et al.* *Science* **281**, 1496–1500 (1998).
2. Fegley, B. Jr & Swindle, T. D. in *Resources of Near-Earth Space* (eds Lewis, J. S., Matthews, M. S. & Guerrieri, M. L.) 367–426 (Univ. Arizona Press, Tucson, 1993).
3. Watson, K., Murray, B. C. & Brown, H. J. *Geophys. Res.* **66**, 3033–3045 (1961).
4. Arnold, J. R. J. *Geophys. Res.* **84**, 5659–5668 (1979).
5. Slade, M. A., Butler, B. J. & Muhleman, D. O. *Science* **258**, 635–640 (1992).
6. Harmon, J. K. & Slade, M. A. *Science* **258**, 640–643 (1992).
7. Harmon, J. K. *et al.* *Nature* **369**, 213–215 (1994).
8. Nozette, S. *et al.* *Science* **274**, 1495–1498 (1996).
9. Stacy, N. J. S., Campbell, C. B. & Ford, P. G. *Science* **276**, 1527–1531 (1997).
10. Binder, A. B. *Science* **281**, 1475–1476 (1998).
11. Lewis, J. S., McKay, D. S. & Clark, B. C. in *Resources of Near-Earth Space* (eds Lewis, J. S., Matthews, M. S. & Guerrieri, M. L.) 3–14 (Univ. Arizona Press, Tucson, 1993).
12. Lewis, J. S. *New York Times* 23 (7 December 1996).

## Daedalus

### Worlds apart

Last week found Daedalus musing on Everett's 'many worlds' interpretation of quantum probabilities. If a quantum choice happens to go one way in our world, there is some other world in which it goes the other way. Thus everything that can possibly happen, does happen, in some world or other. Your wildest secret fantasy is actually coming true somewhere. Conversely, in some world or other, there is an individual for whom your ordinary humdrum life is *his* wildest fantasy.

Indeed, says Daedalus, that is why we have the fantasies we do, or enjoy mental creativity of any kind. The many Everett worlds are embedded in a set of equally valid vector subdivisions with mixed, non-physical quantum states: the 'spirit world', through which information can leak telepathically between the Everett worlds.

But Daedalus now reckons that physical communication may also be possible. A quantum event in this world, such as a radioactive decay, must have its counter-event in another world, to preserve the total sum of possibilities. Daedalus recalls the enigmatic Schmidt machine, whose electric lamps are lit at random by pulses from a radioactive source. Some people, it is claimed, can predict by ESP which lamp is going to light next. Some can even will a given lamp to light next. This telepathic ability to influence quantum states could signal to another Everett world, in which the complementary decay is occurring. Physicists there could read the message in the nonrandom decay pattern of their sample. They could even signal back in the same way.

Now one counting experiment on <sup>60</sup>Co and <sup>137</sup>Cs has found, against all expectation, that their decay was serially nonrandom<sup>1</sup>. Someone, somewhere, is signalling to us! So Daedalus advocates a careful search for the lack of randomness in quantum events. Radioactive decay is an obvious example; electrical circuit noise is another; chaotic phenomena that rapidly amplify sub-Heisenberg irregularities to the macro scale are another. The field is vast and totally neglected; a positive outcome would be revolutionary. Many thinkers have imagined the likely form of coded messages from other civilizations — lists of prime numbers, the digits of  $\pi$ , and so on. They may be taken aback if the first signals seem to be descriptions of other people's fantasies.

David Jones

1. Anderson, J. L. & Spangler, G. W. J. *Phys. Chem.* **77**, 3114–3121 (1973).



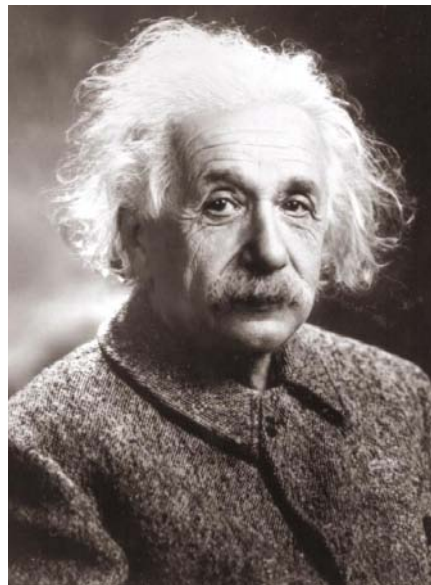
# Icons of intellect

**Herschel the star-gazer with light around his head, Einstein's wild hair and vast brain, Hawking's interstellar mind transcending his earthbound body. It's not just that we've seen them so often: some scientists really look the part.**

V&A PICTURE LIBRARY



Julia Margaret Cameron, "Sir John Herschel", 1867.



Orren Turner, "Albert Einstein", c. 1947, Library of Congress.



James King-Holmes, "Stephen Hawking", 1997.

## Martin Kemp

What a scientist (or artist, author, composer...) looks like should not matter to us. It would make no difference to the theories of relativity if Albert Einstein was clean-shaven, or even if he had a *retroussé* nose. Yet we harbour an apparently irresistible urge to scrutinize the appearance of famous and infamous persons — above all their faces — as if we might use outer signs to understand exceptional inner attributes.

**A select body of images of famous persons have come to serve as more than physiognomic records. They have assumed iconic status, standing for something more archetypal than the traces of individual appearance.**

The fame of the subject is crucial, but the nature of the portrayal must exhibit the potential to work with popular archetypes. Good examples are Yousef Karsh's photograph of Winston Churchill as British 'bulldog', and Andy Warhol's schematic Marilyn Monroes as the all-American blonde bombshell.

Sometimes the image can be of an assumed persona, like Charlie Chaplin's tramp, accompanied by the attributes of hat and stick. Once established, the person can be pictured through minimal cues of caricatured resemblance, particularly if reinforced with simple symbols, like Churchill's cigar.

Few images of scientists have achieved this

status. Leonardo is familiar enough, but he stands for universal genius rather than professional scientist. No face of Sir Isaac Newton readily looms into public view when his name is mentioned. Yet Einstein has come to be instantly recognizable in caricatured form — wild halo of white hair, hooded eyes, characterful nose, bushy moustache and exaggeratedly large 'brain-box' — as seen on T-shirts and orchestrating word-processing 'help' programmes. The theatrically lit image by Orren Turner, photographed at Princeton around 1947, plays as potently on the notion of the venerable sage as any of the photographic studies of the older Einstein.

The general ancestry of Einstein as magus is clear, and more specifically it stands in line of descent from Julia Margaret Cameron's Victorian photographs of the great astronomer and pioneer of photographic processes, Sir John Herschel. So concerned was she to capture the qualities she recognized in the friend whom she called her 'teacher and High Priest', that she cajoled him into washing his hair so that it would encircle his head like an aura of cerebral light.

Such images, once they have assumed canonical status, serve to anchor our mental picture of the celebrity at a particular point in their lives. Although Herschel and Einstein were unquestionably once young, as was Leonardo, it is very difficult to envisage them other than as venerable, wise men, as those who are old enough to 'know the secrets'. There are photographs of Einstein as

an excise clerk, dark haired and photographically naive, but they somehow seem unformed and do not do the iconic job at all.

It would be difficult today to envisage the lighting of a portrait of a great scientist in the same overtly stagy way as those of Herschel and Einstein, but this does not exclude the possibility of new icons. The most likely candidate is Stephen Hawking, Newton's latest successor at Cambridge as Lucasian Professor of Mathematics. His prominence as a public persona, through his authorship of the improbable best-seller, *A Brief History of Time*, and his unembarrassed willingness to disclose the effects of his motor neurone disease, has meant that his appearance has become widely familiar. The contrast between the clear signs of Hawking's physical infirmity, and the public perception of him as a genius of modern science, points up the traditional mind-body duality.

Whereas the portraits of Herschel and Einstein radiated deep wisdom through their external aura, Hawking is, by implication, pure mind trapped in a physical integument whose incapacity mocks the capacity and reach of his intellect. James King-Holmes's 1997 photograph shows how a portrait of Hawking may possess potential not just to portray individual appearance but to embody the kind of general concepts that are necessary for an image to achieve iconic status. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

e-mail: martin.kemp@trinity.oxford.ac.uk

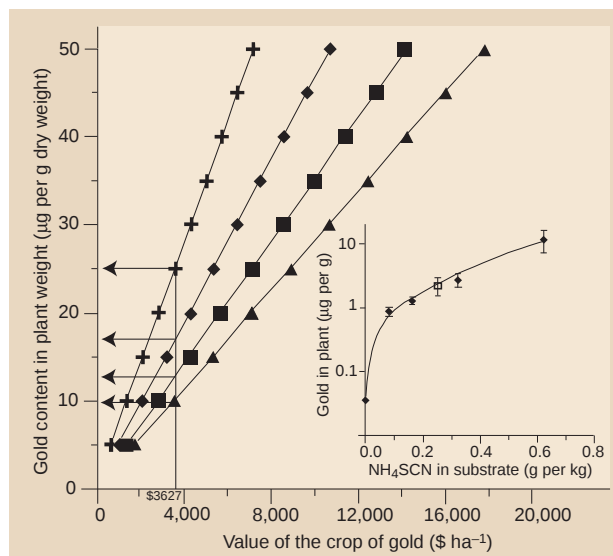


# Harvesting a crop of gold in plants

The possibility of turning base metals into gold has intrigued many scientists since the early alchemists, and the discovery of significant gold uptake by plants has long been a 'philosopher's stone'. But background levels of gold in plants are usually very low, rarely exceeding 10 ng per g dry tissue (10 p.p.b.)<sup>1</sup>. Hyperaccumulator plants<sup>2</sup>, however, have 100 times the elemental concentrations of normal vegetation, a level of 1 mg per g dry tissue (1 p.p.m.). They can be used in phytoremediation<sup>3</sup>, the *in situ* improvement of polluted sites. Hyperaccumulation can be induced by adding a chemical amendment, such as EDTA, to a plant substrate to make soluble an otherwise insoluble target metal, such as lead<sup>4</sup>. Here we have induced plants to accumulate gold from ores by treating the substrate with ammonium thiocyanate. This technique might be used as a form of biological mining (phytomining) for gold<sup>5,6</sup>.

We used ammonium thiocyanate as a substrate amendment because it is commonly used for making gold soluble in mining operations. Table 1 compares the ease of gold extraction in four types of ore. Unweathered sulphide gold ore (not shown in Table 1) from Macraes mine in New Zealand had very little extractable gold, and we were unable to induce plants to remove this to any significant degree. Ore from the Waihi Mine has gold mainly in its native form, which we were able to induce plants to remove. To overcome the problem of sulphide occlusion of gold in some ores and possible non-homogeneity in others, a synthetic finely disseminated colloidal gold ore (made from gold chloride) was prepared and planted with *Brassica juncea*, a plant of high biomass and rapid growth rate. Other experiments using this and other species involved the additional use of a synthetic ore prepared from finely divided gold powder (44 µm) as well as the Tui and Waihi ores crushed to a size of 0.5 mm.

All plants were grown in 250-ml pots containing the appropriate substrate. The



**Figure 1** From gold leaves to gold leaf: the economic value of a gold crop as a function of world prices and concentrations in plant material. Prices: \$200 (crosses), \$300 (diamonds), \$400 (squares), \$500 (triangles). Inset, thiocyanate-induced uptake of gold by *Brassica juncea* from a 5 µg per g finely disseminated synthetic gold ore (diamonds) and natural ore (square).

plants were treated with thiocyanate at rates of 0.00, 0.08, 0.16, 0.32 and 0.62 g per kg dry substrate weight. After seven days, aerial parts of the plants were harvested, dried and analysed for gold by graphite furnace atomic absorption spectroscopy. Gold concentrations in several plants as a function of added thiocyanate are shown in Table 1, and in Fig. 1 (inset) for *Brassica juncea* only.

All plant species showed hyperaccumulation of gold. The highest individual value was 57 µg gold per g dry weight in *B. juncea*. However, values were very variable, perhaps because the higher amendment levels caused necrosis of some plants and the induced gold concentration depended on the time that each plant remained viable. *Brassica* plants grown in the synthetic gold powder substrate (Fig. 1 inset) contained as much gold as those grown in the disseminated substrate. Similar values (9–19 µg per g dry tissue) were obtained for the same species grown in natural Waihi ore.

Induced hyperaccumulation of gold can be used for phytomining, with the resulting auriferous 'bio-ore' sold. The gross value of gold and the chemical costs involved in its extraction using a plant with a biomass of

20 t ha<sup>-1</sup> (ref. 6) are shown in Fig. 1. Assuming that thiocyanate costs \$3 per kg, the figure of \$3,627 represents its total cost per hectare when added to a depth of 15 cm at an application rate of 0.64 g per kg (the highest rate of Fig. 1). As the price of gold increases, the operation becomes more favourable economically. At the current world price of about \$300 per ounce, we would require a gold concentration of around 17 µg per g dry weight in a crop of *B. juncea*. Several values from the experiments shown in Fig. 1 were either above or very close to 17 µg per g. Some of the other phytomining costs may be recouped by selling the energy of plant combustion, as is done in the sugar-cane industry.

Induced hyperaccumulation of gold appears to be relatively independent of plant species, so it should be possible to use plants (such as chicory) that might be easy to grow on mine tailings. Any residual thiocyanate will be broken down rapidly in the substrate<sup>7</sup>.

We believe this is the first evidence of significant gold uptake by any plant. As well as the economic ramifications this technology may have, this ability to make a 'crop of gold' opens up the way for the phytoextraction of other noble metals.

**Christopher W. N. Anderson\*, Robert R. Brooks\*, Robert B. Stewart\*, Robyn Simcock†**

\*Department of Soil Science, Institute of Natural Resources, Massey University, Palmerston North, New Zealand  
e-mail: r.brooks@massey.ac.nz

†Landcare Research New Zealand Ltd, Private Bag, Palmerston North, New Zealand

**Table 1 Conditions and hyperaccumulation data**

| Species                      | Substrate                 | Extraction* (%) | Thiocyanate (g per kg) | Plants | Gold yield (µg per g)† |
|------------------------------|---------------------------|-----------------|------------------------|--------|------------------------|
| <i>Brassica juncea</i>       | Waihi ore‡                | 1.8             | 0.50                   | 4      | 9.27–19.34             |
| Chicory                      | Tui mine§                 | 22.6            | 0.64                   | 5      | 0.07–1.19              |
| <i>Impatiens</i> sp.         | Waihi ore‡                | 1.8             | 0.20                   | 1      | 3.09                   |
| <i>Arrhenatherum elatius</i> | Waihi ore‡                | 1.8             | 0.50                   | 4      | 0.07–1.43              |
| <i>B. juncea</i>             | Disseminated gold in sand | 9.2             | 0.64                   | 12     | 2.13–57.32             |
| <i>B. juncea</i>             | Fine gold powder in sand  | –               | 0.25                   | 8      | 0.37–6.48              |

\*Percentage of gold extracted from 1 g of substrate into 10 ml of 2 g l<sup>-1</sup> ammonium thiocyanate in 24 h.

†Gold values are for dry matter and are whole-plant analyses except for *B. juncea* on fine gold powder 44 (µm) in sand, which are for leaves only.

‡Natural colloidal gold.

§Acid sulphide mine tailings.

1. Dunn, C. E. in *Biological Systems in Mineral Exploration and Processing* (eds Brooks, R. R., Dunn, C. E. & Hall, G. E. M.) 371–426 (Ellis Horwood, Hemel Hempstead, 1995).

2. Brooks, R. R., Lee, J., Reeves, R. D. & Jaffré, T. J. *Geochim. Explor.* **7**, 49–57 (1977).
3. McGrath, S. P., Sidoli, C. M. D., Baker, A. J. M. & Reeves, R. D. in *Integrated Soil and Sediment Research: A Basis for Proper Protection* (eds Eijsackers, H. J. P. & Hamers, T.) 673–676 (Kluwer, Dordrecht, 1993).
4. Robinson, B. H. *et al.* *J. Geochim. Explor.* **60**, 115–126 (1997).
5. Robinson, B. H., Brooks, R. R., Howes, A. W., Kirkman, J. H. & Gregg, P. E. H. *J. Geochim. Explor.* **60**, 115–126 (1997).
6. Blaylock, M. J. *et al.* *Environ. Sci. Technol.* **31**, 860–865 (1997).
7. Huang, C.-H. & Pavlostathis, S. G. *Water Res.* **31**, 2761–2770 (1997).

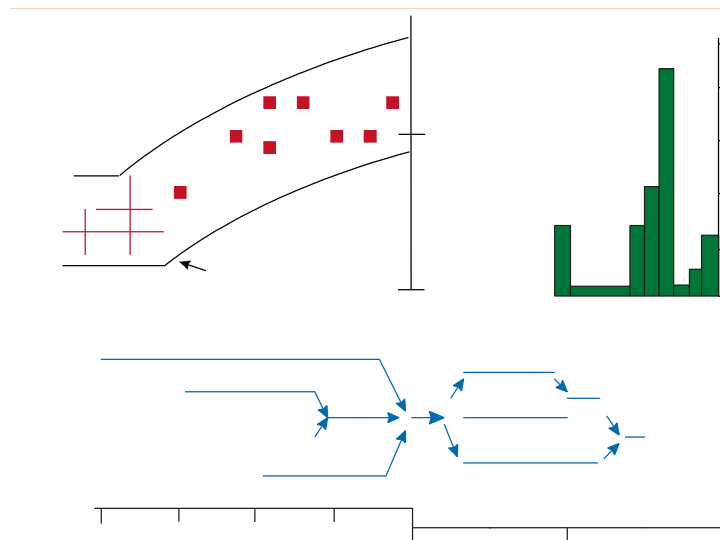
## Salinity history of the Earth's early ocean

It is commonly thought that the oceans are becoming saltier with time as sodium and chlorine are weathered out of continental rocks and transported to the sea. Here we argue that the salinity of the early ocean was 1.5 to 2 times the modern value, and that it did not decline significantly until surprisingly late in the Earth's history. If correct, this theory could help explain why the evolution of higher life took so long.

Volatile compounds such as HCl and H<sub>2</sub>O were probably outgassed from the interior of the Earth early in its history<sup>1</sup>. Chloride ions would have accumulated in the initial ocean with sodium ions leached from surrounding rocks. The only known processes that can remove NaCl from the oceans in significant amounts are the evaporative deposition of salt and the sequestration of brine as deep groundwater on continents. The brine originates from the dissolution of salt and/or as remnants of partly evaporated sea water. Before the continental crust developed, the salt and brine currently found on the continents would have been entirely in the ocean. The oceans became less saline as continental platforms assembled and brine and salt began to accumulate amid a cycle of deposition and erosion within giant sedimentary basins.

If all the known subsurface salt deposits were returned to the sea, salinity would rise by 30% (ref. 1). More chlorine may currently occur in saline groundwater than in salt, perhaps 2–3 times as much<sup>2</sup>. If this chlorine had come from the sea, the initial salinity could have been twice the present level.

There was some continental crust 3.5 billion years ago (Gyr), but large-scale riverine input of <sup>87</sup>Sr derived from granite on emergent continents does not appear in the marine sedimentary record until 2.5 Gyr (Fig. 1). Using a recent model of continental growth<sup>4</sup>, one continental mass existed at 3.0 Gyr and another developed at about 2.5 Gyr (Fig. 1). Two further masses appear at about 2.0 Gyr, from which time the history is dominated by assembly and break-up of continents. If this model is correct,



**Figure 1** History of continents (after ref. 4), record of salt deposition for the past 540 million years (after ref. 2) and Sr isotope record of marine sediments (after ref. 3). The time axis applies to all three panels and is expanded for the past 1.0 Gyr. Salt deposits older than those shown have not been inventoried but probably total less than  $0.5 \times 10^{21}$  g because of recycling and/or non-deposition. The progressive increase in <sup>87</sup>Sr is most readily interpreted as input of riverine Sr from the weathering of continents that emerged largely after 2.5 Gyr.

large-scale deposition of NaCl on continents could not have begun before 2.5 Gyr.

Direct evidence for this highly saline early ocean may exist in chemical data<sup>5</sup> for fluid inclusions in quartz crystals formed in a deep marine hydrothermal environment of 3.2 Gyr. Chloride contents were 165% of the modern value, so the authors inferred that evaporative concentration of sea water had occurred before hydrothermal circulation. An alternative interpretation is that the inclusions are samples of a more saline early ocean.

If large-scale salt deposition and brine storage began at 2.5 Gyr and proceeded at the net rate of the last 540 million years (Fig. 1), the decrease in salinity could have been complete by ~2 Gyr. However, unusual (even improbable) geographic, geologic, climatic, oceanographic and depositional conditions are required for large salt deposits to form. The rate of accumulation was almost certainly lower until continental platforms were widespread at ~2 Gyr. Also, approximately 50% of the known salt was deposited in an interval of 100 million years (Fig. 1). If the conditions that gave rise to this extraordinary accumulation never existed previously, then the decrease in salinity may have persisted until about 1 Gyr, or even to the time of the Cambrian 'explosion' of marine life at 0.54 Gyr.

Most forms of modern macroscopic life cannot tolerate salinities above 50‰ (ref. 6). Cyanobacteria are more salt tolerant than most organisms, and these dominate the Precambrian fossil record. This may be, in part, because higher salinities were an

impediment to the evolution of more complicated life forms. For example, the increase in dissolved oxygen from marine photosynthesizers is commonly invoked to explain the evolution of metazoans<sup>7</sup>. Oxygen solubility in sea water decreases significantly as salinity and temperature increase<sup>8</sup>, so the rise in oxygen level would have been retarded in a more saline ocean. This would be especially true if early Earth temperatures were as high as suggested by isotope data for early sedimentary rocks<sup>9</sup>.

It is commonly assumed that the oceans were the exclusive site of early evolution. However, the lack of fossil constraints, the increasing indications of microbial activity in Precambrian non-marine environments<sup>10</sup> and the inferred higher ocean salinity suggest that this is not necessarily the case. In the present model, the earliest life was either tolerant of salt or was restricted to the more dilute waters of estuaries or entirely non-marine environments. Either way, if continents were not fully developed early on there is a major salinity problem that must be considered when discussing the history of life on Earth or on any object with an overall Cl:H<sub>2</sub>O ratio similar to that of Earth.

**L. Paul Knauth**

Department of Geology,  
Arizona State University,  
Tempe, Arizona 85287-1404, USA  
e-mail: knauth@asu.edu

1. Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* (Princeton Univ. Press, 1984).
2. Land, L. S. *Aquat. Geochim.* **1**, 137–145 (1995).
3. Mirota, M. D. & Veizer, J. *Geochim. Cosmochim. Acta* **58**, 1735–1745 (1994).

4. Rogers J. J. W. *J. Geol.* **104**, 91–107 (1996).
5. DeRonde, C. E. J., Channer, D. M., DeR., Faure, K., Bray, C. J. & Spooner, E. T. C. *Geochim. Cosmochim. Acta* **61**, 4025–4042 (1997).
6. Hughes Clarke, M. W. & Keij, A. J. in *The Persian Gulf: Holocene Carbonate Sedimentation and Diagenesis in a Shallow Epicontinental Sea* (ed. Purser, B. H.) 33–56 (Springer, Berlin, 1973).
7. Knoll, A. H. *Science* **256**, 622–627 (1992).
8. Weiss, R. F. *Deep Sea Res.* **17**, 721–735 (1970).
9. Knauth, L. P. & Lowe D. R. *Earth Planet. Sci. Lett.* **41**, 209–222 (1978).
10. Horodyski, R. J. & Knauth, L. P. *Science* **263**, 494–498 (1994).

## Receptor that leaves a sour taste in the mouth

The ability to detect taste stimuli results from the activation of taste receptors located in taste-bud cells. There are several gustatory transduction mechanisms, involving membrane receptors, guanine-nucleotide-binding proteins (G proteins), second messengers and ion channels<sup>1</sup>, but genes encoding taste receptors have not yet been identified. Here we identify a complementary DNA that encodes a receptor for sour tastes.

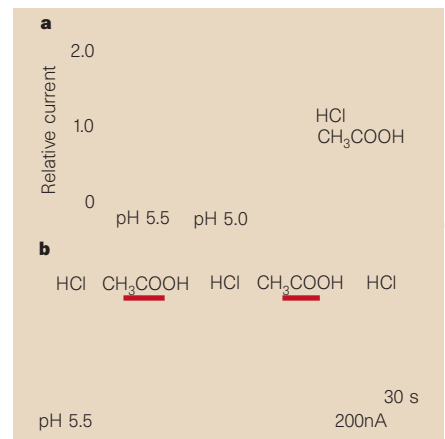
There are five basic tastes: sour, salty, bitter, sweet and umami (the taste of glutamate and similar naturally occurring substances)<sup>1</sup>. Sour-tasting substances, evoked by acids, interact directly with ion channels in taste-receptor cells<sup>1</sup>. In rats, amiloride-sensitive cation channels may directly mediate part of sour-taste transduction<sup>2</sup>.

To isolate a sour-taste receptor, we

constructed a rat circumvallate papilla cDNA library and screened it using combined homology and functional expression approaches. We screened the library at low stringency with oligonucleotide probes encoding proteins that shared significant homology with the second putative transmembrane domain of the amiloride-sensitive/degenerin cation-channel family. We obtained 18 independent cDNA clones, restriction-enzyme analysis of which produced several digestion patterns.

To find candidate sour-taste receptors, we expressed these clones in *Xenopus laevis* oocytes and analysed them using a two-electrode voltage clamp. Injecting a mixture of 5'-capped complementary RNA transcripts from all positive clones into *Xenopus* oocytes resulted in no amiloride-sensitive current at pH 7.5, but altering the extracellular pH to 5.5 produced a large inward current, which could be partly blocked by amiloride. We detected almost no current in water-injected oocytes with the same pH alterations.

Each clone from the pool was then transcribed independently and the cRNA products were injected separately into *Xenopus* oocytes. Only one clone showed the characteristics of the proton-gated amiloride-sensitive cation channel (data not shown). Simultaneous *in situ* hybridization showed that this clone was expressed in the taste buds in the circumvallate papilla of the rat tongue. Sequence analysis showed that this clone was identical to mammalian



**Figure 2** Expression of MDEG1 in *Xenopus* oocytes. **a**, Whole-cell currents of oocytes expressing MDEG1 at  $-70$  mV. Data are plotted relative to the current elicited by ND140 solution adjusted with HCl. Values are mean  $\pm$  s.e.m. from seven different oocytes. **b**, Representative current trace of an oocyte injected with cRNA encoding MDEG1. ND140 solution (containing (in mM) 140 NaCl, 2 KCl, 1.8  $\text{CaCl}_2$ , 1  $\text{MgCl}_2$  and 5 HEPES, pH 5.5) adjusted with HCl or  $\text{CH}_3\text{COOH}$  was applied alternately.

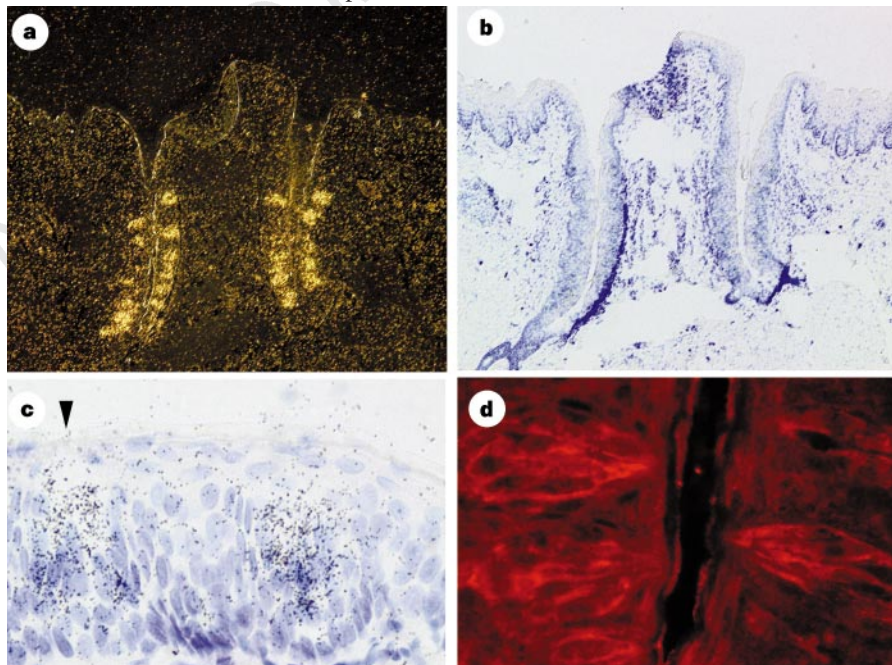
degenerin-1 (MDEG1)<sup>3–5</sup>, although minor nucleotide differences were found in non-coding regions.

We localized MDEG1 in the taste buds of the circumvallate papillae by using *in situ* hybridization. There were specific hybridization signals for MDEG1 messenger RNA in most of the taste buds on both sides of deep clefts in the rat circumvallate papilla, but none in any surrounding tissue (Fig. 1a, b). Under high magnification, the dense hybridization signals are seen to accumulate only in the central portion of the taste buds, indicating that this mRNA was not expressed in the basal cells (Fig. 1c). A ribonuclease protection assay confirmed that MDEG1 mRNA is expressed in circumvallate papillae (data not shown). Thus, MDEG1 mRNA is expressed in the taste buds of the rat circumvallate papilla.

We used immunohistochemistry to determine the subcellular localization of MDEG1 protein in the taste buds. We prepared a polyclonal antiserum against MDEG1 by immunizing rabbits. MDEG1 immunoreactive cells occurred only in the taste buds, not in surrounding tissues.

At the subcellular level, intense MDEG1 immunoreactivity was seen at the plasma membrane and the apical part of the immunoreactive taste-bud cells (Fig. 1d). Immunoelectron microscopic analysis showed that MDEG1-positive taste-bud cells had the characteristics of type III cells (primary taste-receptor cells) (data not shown)<sup>1,6,7</sup>.

Protons cause acidic taste, and they activate sour-taste receptors<sup>1</sup>. However, acetic acid, the main ingredient of vinegar, is more sour than hydrochloric acid at equal



**Figure 1** Expression and localization of MDEG1. **a–c**, Expression of MDEG1 mRNA in the rat circumvallate papilla by *in situ* hybridization. **a**, **b**, The circumvallate papilla, stained with an antisense probe, in dark-field (**a**) and bright-field (**b**) photomicrographs. **c**, Taste buds stained with an antisense probe, under high magnification. The dense hybridization signals for MDEG1 mRNA accumulate only in the central portion of the taste buds. The arrowhead indicates a taste pore. **d**, Localization of MDEG1 by immunohistochemistry in the circumvallate taste buds. Intense MDEG1 immunoreactivity was seen at the plasma membrane and the apical part of the MDEG1-positive cells in the circumvallate taste buds.



pH. In an attempt to explain this phenomenon, we applied ND140 solutions adjusted to pH 5.5 with acetic acid or hydrochloric acid to oocytes injected with cRNAs encoding MDEG1. Stimulation by acetic acid generated larger inward currents than those induced by hydrochloric acid at equal pH (Fig. 2a, b). These response patterns of MDEG1 to the two acids are similar to characteristics of sour-taste sensation of the acids *in vivo*.

Our molecular biological, morphological and electrophysiological analyses indicate that MDEG1 acts as a receptor for sour tastes in taste-bud cells.

Shinya Ugawa\*†, Yuki Minami†, Wei Guo\*†, Yoshitsugu Saishin\*†, Koichi Takatsuji†, Takashi Yamamoto‡, Masaya Tohyama†, Shoichi Shimada\*

\*Department of Anatomy II,  
Nagoya City University Medical School,  
1 Kawasumi, Mizuho-cho, Mizuho-ku,  
Nagoya 467-8601, Japan

†Department of Anatomy and Neuroscience,  
Osaka University Medical School,  
2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

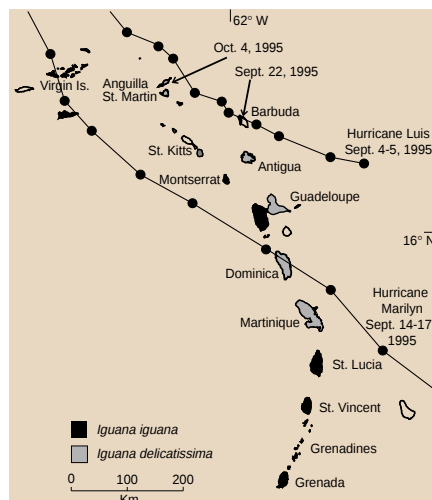
‡Department of Behavioural Physiology,  
Faculty of Human Sciences, Osaka University,  
1-2 Yamadaoka, Suita, Osaka 565-0871, Japan  
e-mail: ugawa@med.nagoya-cu.ac.jp

1. Lindemann, B. *Physiol. Rev.* **76**, 719–766 (1996).
2. Harris, D. E. *et al. Chem. Senses* **19**, 481–482 (1994).
3. Waldmann, R. *et al. J. Biol. Chem.* **271**, 10433–10436 (1996).
4. Bassilana, F. *et al. J. Biol. Chem.* **272**, 28819–28822 (1997).
5. Lingueglia, E. *et al. J. Biol. Chem.* **272**, 29778–29783 (1997).
6. Murray, R. G., Murray, A. & Fujimoto, S. *J. Ultrastruct. Res.* **27**, 444–461 (1969).
7. Fujimoto, S. & Murray, R. G. *Anat. Rec.* **168**, 383–413 (1970).

## Over-water dispersal of lizards due to hurricanes

The possibility and probability of over-water dispersal as a mechanism to explain the distribution of terrestrial animal species in the Caribbean has been hotly debated since the early part of this century<sup>1,2</sup>. Each theory that has been proposed — including land bridges and over-water dispersal — has involved over-water dispersal to some extent in the distribution of animals. Yet many people remain sceptical of over-water dispersal, believing that the use of rafts is improbable, unobservable and consequently untenable. Here we present evidence to support over-water dispersal as the mechanism by which green iguanas colonized Anguilla.

For over-water dispersal to be considered a realistic explanation for the distribution of species in the Caribbean, it must be demonstrated that a viable population could be established. This can be accomplished by the invasion of either a pregnant individual, an asexually reproducing



**Figure 1** Tracks of hurricane Luis and hurricane Marilyn through the islands of the Lesser Antilles in the Caribbean. Dates with arrows indicate the first sightings of iguanas.

individual or several individuals of both sexes. Animals have been observed on rafting flotsam<sup>3</sup>, but most of these are small organisms, such as insects<sup>4,5</sup>. The few accounts of vertebrates found on rafts<sup>6,7</sup> have reported only single individuals (but see ref. 8), and do not provide convincing evidence that a population can become established once landfall is reached.

On 4 October 1995, at least 15 individuals of the green iguana, *Iguana iguana*, appeared on the eastern beaches of Anguilla in the Caribbean. This species did not previously occur on the island. They arrived on a mat of logs and uprooted trees, some of which were more than 30 feet long and had large root masses. Local fishermen say the mat was extensive and took two days to pile up on shore. They reported seeing iguanas on both the beach and on logs in the bay.

Dispersal events are often assumed to be caused by large storms, such as hurricanes<sup>9</sup>. The 1995 hurricane season had above normal activity, with 11 hurricanes and 8 tropical storms<sup>10</sup>. On 4 and 5 September 1995, hurricane Luis moved through the eastern Caribbean (Fig. 1). The storm was rated category 4 on the Saffir/Simpson Hurricane Scale (SSHS). Hurricane-force winds were reported as far south as Guadeloupe. A week and half later, on 14–17 September, hurricane Marilyn (SSHS category 2) followed a parallel path slightly south of hurricane Luis, and many of the same islands once again experienced hurricane-force winds (Fig. 1). Approximately a month after the first of these hurricanes, iguanas reached the shores of Anguilla.

A survey was established within a month of the invasion. We preserved one individual (CM 145848) and marked seven others. Another iguana escaped before processing, and three more were caught by the fishermen but died shortly afterwards. A further

three iguanas were sighted on Scrub Island, 0.5 km northeast of Anguilla, and green iguanas were also reported on the northeast coast of Barbuda, 150 km southeast of Anguilla (D. V. Nicholson, personal communication).

Captured individuals (three males and five females) ranged in snout–vent length from 276 to 400 mm. Iguanas were captured between December 1995 and March 1998. Our most recent sighting of *I. iguana* on Anguilla, on 11 March 1998 (29 months after the invasion), was of a female with enlarged ovarian follicles (possibly oviductal eggs). Because both males and females invaded the island, survived and appear to be healthy, with a female in reproductive condition, the likelihood of reproduction is high.

There are two species of *Iguana* in the Lesser Antilles: *I. iguana* and *I. delicatissima* (Fig. 1). *I. iguana* is widespread, occurring from Mexico through Central America into South America as far south as northern Paraguay, and is also found on many of the southern islands of the Lesser Antilles. *I. delicatissima* occurs on most islands in the Lesser Antilles where *I. iguana* is absent. The track of hurricanes Luis and Marilyn (Fig. 1), the current distribution of *I. iguana* in the Lesser Antilles (Fig. 1), and the general west-northwest ocean currents in the region suggest that these iguanas originated on the island of Guadeloupe.

The probability that a species will successfully colonize an island depends on both the probability that it will reach the island and the probability that it will survive once it arrives. Our observations confirm that raft dispersal can occur successfully, and document the over-water dispersal of a group of large vertebrates and their persistence and possible reproduction after landfall.

**Ellen J. Censky**

Section of Amphibians and Reptiles,  
Carnegie Museum of Natural History,  
Pittsburgh, Pennsylvania 15213, USA  
e-mail: censky@clpgh.org

**Karim Hodge**

Anguilla National Trust,  
Anguilla, British West Indies

**Judy Dudley**

National Council of the Paper Industry for Air and  
Stream Improvement,  
West Coast Aquatic Biology Research Laboratory,  
Anacortes, Washington 98221, USA

1. Williams, E. E. in *Biogeography of the West Indies, Past, Present, and Future* (ed Woods, C. A.) 1–46 (Sandhill Crane, Gainesville, 1989).
2. Crother, B. I. & Guyer, C. *Herpetologica* **52**, 440–465 (1996).
3. Heatwole, H. & Levins, R. *Ecology* **53**, 112–117 (1972).
4. Wheeler, W. M. *Psyche* **23**, 180–183 (1916).
5. Holzapfel, E. P. & Harrell, J. C. *Pacif. Insects* **10**, 115–153 (1968).
6. King, W. *Quart. J. Florida Acad. Sci.* **25**, 45–52 (1962).
7. Hardy, J. D. *Bull. Maryland Herpetol. Soc.* **18**, 37–142 (1982).
8. Boyd, C. E. *Herpetologica* **18**, 269 (1962).
9. Darlington, P. J. *Quart. Rev. Biol.* **13**, 273–300 (1938).
10. <http://www.nhc.noaa.gov/pastall.html>

# Calls for a ceasefire in the science wars

**Making Natural Knowledge: Constructivism and the History of Science**

by Jan Golinski

Cambridge University Press: 1998. 236pp.  
£37.50, \$54.95 (hbk); £13.95, \$16.95 (pbk)

John Henry

Constructivism, as it is used here, refers to the “central notion that scientific knowledge is a human creation, made with available material and cultural resources, rather than simply the revelation of a natural order that is pre-given and independent of human action”. Accordingly, it is based on the work of recent philosophers and, especially, sociologists of science who have sought to extend our understanding of scientific knowledge (from its first gleaming to its consolidation as the scientific consensus) by paying close attention to the complex ways in which scientists interact with one another and with other aspects of their social milieu.

The introduction and first chapter of this book provide a clear and illuminating brief

account of this very recent tradition of the ‘sociology of scientific knowledge’. It begins with the work of Thomas Kuhn and the way that this work was interpreted by the ‘Edinburgh school’ to develop the ‘strong programme’ in the sociology of scientific knowledge, and ends with an account of the more recent claims of the French sociologist Bruno Latour. The subsequent chapters are rather different, showing some of the ways in which historians of science might be said to have deployed ideas from the sociology of scientific knowledge to inform their historical analyses of the development of science. These later chapters, therefore, constitute what Jan Golinski describes as “an extended historiographical essay, a review of recent writing about the history of the sciences”.

Although Golinski does not include practising scientists in the list of academic readers he feels will benefit from this book (it is primarily intended for graduate students in the history of science, and is in fact the latest addition to a distinguished series from Cambridge University Press, simply entitled

“Cambridge History of Science”), it may be of interest to those who have been following what has come to be called the ‘science wars’. Those who have so far only troubled to read the contributions to the debate by scientists may be surprised to read Golinski’s entirely standard insistence that social constructivism “should *not* be taken to imply the claim that science can be entirely reduced to the social or linguistic level, still less that it is a kind of collective delusion with no relation to material reality”. Social constructivist analysts of science continue to be astonished that scientists (together with various kinds of philosophers of science and other followers) seem to think that claims about the sociological content of knowledge are somehow exclusive of all other claims. Some otherwise seemingly intelligent commentators seem to think that scientific knowledge is either based on the real world, or based on social interactions between scientists (who, allegedly, collectively agree on what story about the physical world they are going to tell laypeople). This is foolish in the extreme.

Consider, for example, the case of Johannes Kepler’s laws of planetary motion. Although Kepler hit upon his laws by repeated trial-and-error calculations based on the extremely accurate planetary observations of Tycho Brahe, no less a contemporary than Galileo dismissed his efforts on the grounds that they were based on “puerilities” such as action-at-a-distance, cosmic harmonies, and the like. There was a long social process before Kepler’s laws of planetary motion came to be accepted as accurate descriptive representations of the way planets move.

The full historical account of this process would have to take into consideration not only Kepler’s idiosyncratic religious beliefs, and his magical Neoplatonism, which for him were intimately bound up with the laws he had discovered, but also the attitudes to those beliefs of contemporary astronomers and natural philosophers. It would also have to take account of attitudes to Copernicanism after its condemnation by the Roman Catholic church, as well as more technical considerations such as how the prevailing Aristotelian theories of motion could be refined or replaced in order to account for the incessant motions of a body such as the Earth that was clearly not made of some ethereal fifth essence.

Refinements to Kepler’s second law (which claimed that the radius vector from the Sun to a given planet sweeps out equal areas in equal times) were suggested by Ismael Boulliau and Seth Ward, and for a while seemed preferable to Kepler’s version, even though they were less mathematically accurate. Their popularity can be fully

Some of us are  
looking at the stars:  
Tycho Brahe in his  
observatory in 1587.



HULTON GETTY



explained only by studying the expectations of contemporary astronomers as to what was required in astronomical explanations of planetary movements. Putting it crudely, Kepler's version seemed to imply that the planet had to have some means of varying its own speed in order to ensure that it swept out the correct area.

The full story of how Kepler's laws came to be established as legitimate scientific knowledge is immensely rich and complex (and fascinating). But nowhere in the full story is anything revealed which shows that Tycho Brahe's observations were not unprecedentedly accurate, or that Kepler's literally painstaking calculations were hopelessly compromised by, say, his religious beliefs — much less that there are no planets out there to be observed, and that even if there were they would be moving any which way. Sociological accounts of scientific knowledge are complementary to the scientific accounts of that knowledge provided by scientists; they are not futile attempts to provide complete alternatives to it.

The bulk of Golinski's book, after the opening theoretical sections, is devoted to summary accounts of various historical case studies of science that are supposed to exemplify various broad themes from sociology of scientific knowledge. These include studies of the formation of the identity of the scientific practitioner; the historical role of scientific institutions and new disciplinary structures; a consideration of the locations in which science is created, whether in the laboratory or in the field; the place of rhetoric in science; and the role of instruments and other physical apparatus in the formation of scientific knowledge.

Although much of this material is interesting in its own right, I doubt that Golinski would see his summaries, incisive and enlightening though they are, as an adequate substitute for reading the original historical studies upon which he has drawn. The book therefore works well as a textbook intended for students of the history of science, who might be expected to use it as an introduction to further reading, but I imagine it will prove less satisfactory to scientists wishing to learn about the history of science. Indeed, there is a final coda to the book, "The Obligations of Narrative", which is entirely devoted to considerations about the best way for historians to write the history of science.

Another aspect of the book that may diminish its interest for followers of the science wars is the fact that, despite its opening concern with the theoretical basis for the sociology of scientific knowledge, the relevance of much of the subsequent historical material to scientific knowledge is by no means always clear. When Golinski recounts Mario Biagioli's exciting work on Galileo's attempts to secure patronage from the Medici court, for example, he fails to bring

out the relevance of this to the intellectual content of Galileo's scientific work. Biagioli's account of the way Galileo sought to gain and maintain Medicean patronage is the only one I know (and, believe me, the scholarship on Galileo is of industrial proportions) which explains why such a supreme anti-Aristotelian as Galileo promoted an essentially Aristotelian account of comets (as atmospheric phenomena) in his *Assayer* of 1623. All other commentators on Galileo hurriedly gloss over this as an embarrassing aberration by the great man; only Biagioli provides a plausible reason for it. The account is largely sociological: Biagioli does not dwell on the observational or theoretical reasons Galileo had for dismissing Jesuit claims that comets were beyond the sphere of the Moon, but there certainly were reasons of that kind.

The point is, according to sociology of scientific knowledge, that social context and scientific content should not be separated; they are sides of the same coin. Golinski misses a trick, however, by focusing on Biagioli's more straightforwardly historical claims about Galileo's "self-fashioning" as a court philosopher to Grand Duke Cosimo II de Medici. This is just one example, but generally speaking I think it is fair to say that Golinski's summaries of historical work tend to concentrate more on the details of the historical context than on the relevance of this background to scientific thinking. □

John Henry is in the Science Studies Unit, University of Edinburgh, 21 Buccleuch Place, Edinburgh EH8 9LN, UK.

## From fins to limbs and back again

At the Water's Edge: Macroevolution and the Transformation of Life by Carl Zimmer.  
Free Press: 1998. 290pp. \$25, £17.99  
Michael S. Y. Lee

Major evolutionary transitions are one of the most fascinating topics in biology, and have been investigated for over a century by some of the most gifted scientists that ever lived. Vertebrates, as the group that includes ourselves, are the most conspicuous and well-known animals, so it is surprising that the origins of many vertebrate higher taxa — tetrapods, dinosaurs, birds, snakes, turtles, whales and ichthyosaurs, for instance — remained contentious and poorly understood until very recently. But after more than a century of frustratingly slow progress, the past decade has seen a series of fortuitous fossil discoveries, which, coupled with increasingly rigorous analytic methods, have allowed us to understand in great detail the early evolution of these groups.

Carl Zimmer's book focuses on recent advances in two of these areas. The first half deals with the transition from fish to limbed vertebrate, and the second half considers that from hoofed mammal to whale. In both instances, discoveries of striking transitional



## Now you see them, now you don't

Camouflage can't protect the Northern red salamander from the main threat to its life. Amphibians are highly sensitive to disturbance in their habitat, so — like canaries in a coal mine — their declining populations around the world are a clear indicator of environmental damage.

Meanwhile, in *Salamanders of the United States and Canada* (Smithsonian Institution Press, \$60/£46.95) James W. Petranka tells everything you could ask about them in a clear format with nearly 500 photos. It is the first comprehensive survey of the continent's 127 species since 1943.



fossils by palaeontologists have attracted the attention of functional morphologists, systematists and molecular and developmental biologists. The result is modern biology at its most lively and successful, with scientists from diverse disciplines joining forces to probe major evolutionary problems.

There are two difficulties in attempting to review such work. The first is that, with the field moving so rapidly, the book is in grave danger of quickly becoming obsolete. For instance, *Taking Wing*, the recent book by Pat Shipman on the dinosaur–bird transition, was inevitably dated even before it appeared. Similarly, many of the ideas discussed approvingly in Zimmer's book are already being questioned: work on *Ichthyostega* suggests that the early evolution of the tetrapod ear was more complex than previously thought; the bifurcating model of limb development seems now to be an oversimplification; and the molecular evidence for close affinities between baleen and sperm whales has been refuted. However, Zimmer has done everything possible to ensure the

book has a long shelf life: even articles from 1997 have been fully integrated into the text. Unless many more advances are made in the near future, the information will remain relevant for some time.

*At the Water's Edge* is intended to be read from cover to cover by any educated layperson, and this leads to a second obstacle. The relevant research radiates across fields as seemingly disparate as fossils and *Hox* genes, making it difficult to summarize and weave into a coherent, linear story. But the author meets the challenge admirably. Although having primarily an arts background, he clearly understands the diverse scientific issues involved, and cuts through the scientific jargon so anyone can comprehend them. Also, he carefully chooses when to introduce each topic, almost as if unravelling a plot. For instance, in the section on tetrapod origins, he discusses the early palaeontological ideas about the origin of the tetrapod limb, then how these ideas were challenged by recent fossil discoveries, followed by the effect of these discoveries on

models of limb development, and finally how developmental molecular genetics might in turn shed light on the fundamental mechanistic factors underlying the fin–limb transition. Although the origins of land vertebrates and whales might seem to have little in common, the book draws them together by stressing that they exemplify important and general macroevolutionary principles, such as contingency, exaptation and correlated progression.

There are occasional digressions which should have been trimmed, though. In the final chapter, for example, the smooth flow is interrupted by a short (and thus inevitably oversimplified) discussion of punctuated equilibrium, which has little relevance to anything else in the book. This idea is only testable in organisms with a dense fossil record, and (as Zimmer acknowledges) the remains of early tetrapods and whales are much too rare to infer whether they evolved in a gradualist or punctuationist pattern.

Zimmer's book will certainly help an important area of biological research garner the public attention it deserves. But biologists eager to extract the new scientific information in it quickly may become impatient at the sections in which elementary biological principles are explicated at length, or personalities described in lively detail. Nevertheless, all evolutionary biologists should read it thoroughly: it covers advances in so many diverse disciplines that even researchers at the cutting edge of one of these will discover much that is new and relevant in other fields. □

Michael S. Y. Lee is in the Department of Biological Sciences, Monash University, Clayton, Victoria 3168, Australia.

## Winning by a whisker

**Electronic Genie: The Tangled History of Silicon**  
by Frederick Seitz and Norman Einspruch  
University of Illinois Press: 1998. 281pp.  
\$34.95

James D. Meindl

"In truth, one might say that this book has had at least a hundred authors." This enlightening comment from the preface of *Electronic Genie* reveals its strongest feature — the worldwide network of highly knowledgeable advisers and contributors that aided its thorough search into "the tangled history of silicon".

No one was better qualified to lead such a search than Frederick Seitz, one of the founders of materials science as a discipline. His extraordinarily long career has spanned much of the period covered by this book:



**The cutting of Silicon Valley:** clockwise from left, a technician is dwarfed by Baby, the first programmable computer, built in 1948 and recently restored; office staff in 1956 would have needed weight-training to move the world's first portable computer, Baby-Brain E101; by 1992, the tiny Intel Pentium microprocessor was the most complex ever developed.

his pioneering *Modern Theory of Solids* (McGraw-Hill) was published in 1940.

Critical historical questions that the authors set out to answer include: How did crystalline silicon become available for very early wireless telegraphy experiments? Why did the Bell Telephone Laboratories focus on the development of semiconductor triodes or transistors immediately after the Second World War? How did it happen that the invention of the integrated circuit occurred in two start-up companies? *Electronic Genie* answers these queries, and more broadly describes with authority the historical course of development of silicon technology, the principal driver of the modern information revolution.

The book begins with the discovery of the element silicon in the early nineteenth century and concludes with a capsule summary of future prospects. The development of silicon technology is attributed to both the push of fundamental advances in materials science, and the pull of communications, radar and computer systems requirements. Perhaps the most fascinating pathway that is traced begins with the introduction, near the turn of the century, of metallurgical grade silicon. It was originally refined for use in iron-silicon alloys or steel, for such applications as electric power transformers. In electronics it was first used to fabricate rectifying diodes for early twentieth-century wireless telegraphy receivers.

Interest in silicon for electronic devices remained relatively dormant until the early stages of the Second World War, when it was recognized that a chip of silicon in contact with a fine rod of tungsten — a combination known as a 'cat whisker' — provided the critical crystal rectifier needed for effective return signal detection in ground-based and airborne radar receivers. Radar systems using silicon diode detectors proved to be of enormous advantage to the Allies.

In 1948, the point contact bipolar transistor, which uses two proximate cat whiskers forming electrical point contacts on a chip of semiconductor material, was invented at the Bell Telephone Laboratories. In short order, the PN junction was discovered in an ingot of highly purified zone-refined silicon; aligned point contacts on opposite sides of a thin layer of germanium demonstrated a 'two-sided' point contact bipolar transistor; and the bipolar junction transistor was invented, marking the birth of modern microelectronics.

A decade later the invention of the integrated circuit enabled the interconnection of several transistors to form a functional circuit within a single silicon microchip. This momentous achievement presented the opportunity to take advantage of the operating speed, power, size, reliability and cost advantages of the transistor relative to earlier vacuum-tube electronics.

*Electronic Genie* is a resounding success in providing a highly informative and stimulating account of this story. □

James D. Meindl is in the Microelectronics Research Center, Georgia Institute of Technology, 791 Atlantic Drive NW, Atlanta, Georgia 30331, USA.

## The mother of all questions

### The Fifth Miracle: The Search for the Origin of Life

by Paul Davies

Allen Lane: 1998. 260pp. £18.99

Stefan Bengtson

It may be that *The Fifth Miracle* is a misnomer, and the creation of life was only the Third Miracle. (Biblical scholars told Paul Davies that the creation of the Universe, as recounted in Genesis, wasn't really a separate miracle; and shouldn't we just write off the creation of dry land as a mere mopping-up act — leaving the creation of light and of the firmament as the first two?) The three miracles might then correspond to the classic subdivisions of science: physics, chemistry and biology. That these are deeply entangled we already know. If we can figure out in what way they are entangled, we will understand something fundamental about life, the Universe and everything.

Davies's book is a small miracle in itself. In a little more than 200 pages he pursues the Mother of All Questions — "What is life?" — in a way that should be deeply satisfying to physicists, chemists and biologists alike, and he does this in a clear and potent style that should make the book equally stimulating to non-scientists. (Whoever says that you cannot write about science both simply and accurately hasn't read Paul Davies.) A few slips-of-the-keyboard remind us that the writer is not an expert in natural history (or Nordic languages), but this professed "simple-minded physicist" still demonstrates a better grasp of the complexity and uniqueness of living systems than most scientists, biologists included.

The cover shows Earth and Mars in close apposition. Too close, to be exact, but this is to illustrate one of Davies's ideas, that Earth and Mars are not quarantined and never

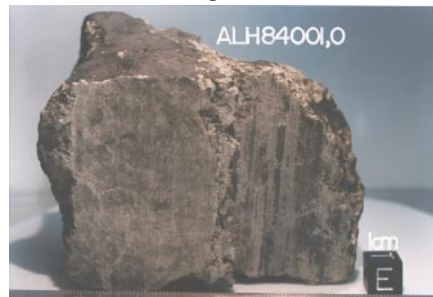
have been. Rocks travel from Mars to Earth (one of them killed a dog!) and probably in the other direction too, and microorganisms thrive in rocks, as we have lately become aware. Some rocks travel the distance in only a few thousand years, so the possibility of cross-contamination with hardy microorganisms is considerable. This is fascinating in itself, but whether or not it means that we are all Martians or that the Earth's biosphere extends (or has extended) to the asteroid belt or beyond, it makes the issue of life on other planets in the Solar System almost trivial.

There is then the greater question, whether or not we are alone in the Universe. Davies has little patience with those who take for granted that as soon as there are Earth-like conditions, life will sprout. This may be true, he says, but then we're making a gigantic assumption that should not be taken lightly. To explore this assumption, he takes the reader on an intellectual joyride up and down the entropy slope, along the way pointing his sharp flashlight at often murky concepts, such as order, organization, chance, randomness, specificity, language and semantics.

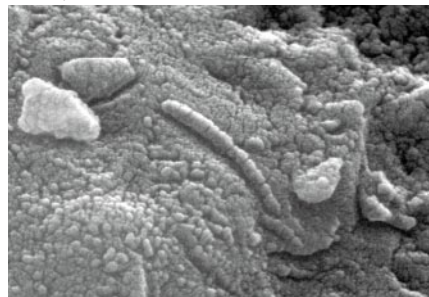
The question of where the information content of life ultimately comes from is, of course, not answered fully, but Davies speculates that gravitation plays a crucial role by particularizing otherwise uniform matter, and that the famous wave-particle duality of quantum mechanics reflects a software-hardware entanglement built into the Universe itself. Biological order may then be ascribed to emergent properties of complex systems, with Darwinian selection serving to distil information from the environment. Natural selection adds information by removing possibilities.

Is life then also the final miracle? Hardly. Just as physics does not fully explain chemistry and chemistry does not fully explain biology, so life does not fully explain consciousness, and consciousness does not fully explain human culture. Davies dwells only briefly on mind and consciousness, but his book is a wonderful example of science as an expression of human culture, and so it truly belongs to the Fifth Miracle. □

Stefan Bengtson is in the Department of Palaeozoology, Swedish Museum of Natural History, Box 50007, SE-104 05 Stockholm, Sweden.



Rock folly: bug-like forms, above right, one-thousandth the width of a human hair, once raised hopes that life existed on Mars before the 4.5 billion-year-old rock, above left, fell to Earth.



NASA



# A genetic framework for floral patterning

François Parcy<sup>†</sup>, Ove Nilsson<sup>†\*</sup>, Maximilian A. Busch<sup>\*</sup>, Ilha Lee<sup>†\*</sup> & Detlef Weigel<sup>\*</sup>

<sup>\*</sup> The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

**The initial steps of flower development involve two classes of consecutively acting regulatory genes. Meristem-identity genes, which act early to control the initiation of flowers, are expressed throughout the incipient floral primordium. Homeotic genes, which act later to specify the identity of individual floral organs, are expressed in distinct domains within the flower. The link between the two classes of genes has remained unknown so far. Here we show that the meristem-identity gene *LEAFY* has a role in controlling homeotic genes that is separable from its role in specifying floral fate. On the basis of our observation that *LEAFY* activates different homeotic genes through distinct mechanisms, we propose a genetic framework for the control of floral patterning.**

The development of multicellular organisms requires the repeated generation of complex tissue and organ patterns from fields of undifferentiated cells. A genetically tractable model for pattern formation in plants is the morphogenesis of individual flowers: a collection of undifferentiated cells termed the floral meristem produces four types of organ in a stereotypic fashion. A landmark in developmental genetics was the proposal of the ABC model, which describes how three classes of homeotic genes act in discrete domains to specify the identity of floral organ types<sup>1,2</sup> (Fig. 1a). Subsequent molecular analysis showed that region-specific activity of the ABC genes is largely regulated at the transcriptional level<sup>3</sup>.

Little is known about how the initial pattern of ABC gene expression is generated. As region-specific expression of the *Arabidopsis thaliana* A-function gene *APETALA1* (*API*) is merely a consequence of repression by the C-function gene *AGAMOUS* (*AG*)<sup>4</sup>, a minimal model for the generation of the ABC pattern needs to explain how *API* is initially activated throughout the flower, and how *AG*, as well as the B-function genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), are activated at a later time point in their specific domains.

Candidates for upstream regulators of flower-specific ABC genes are early-acting genes such as *LEAFY* (*LFY*). Unfortunately, it has been difficult to determine the directness of the interaction between *LFY* and ABC genes<sup>5–7</sup>, because *LFY* affects the identity of the flower meristem itself, an event that precedes the activation of ABC genes. *lfy* null mutations cause a transformation of the first few flowers into leaves with associated shoots which do not express any of the ABC genes simply because these structures never acquired any floral identity<sup>8–10</sup>. Later-arising flowers are replaced by leaf-like bracts in *lfy* mutants, and abnormal flowers develop from the base of the bracts. In these flowers, *AP3* and *PI* expression is very much reduced; *AG* expression is delayed; and *API* expression is almost normal<sup>7,11</sup>. Again, it is difficult to decide whether these alterations are a consequence of missing *LFY* activity at the time that these genes are activated, or whether they result from an earlier defect in specification of the flower meristem. Moreover, expression of ABC genes in the abnormal flowers of *lfy* mutants indicates that *LFY* is not absolutely required for their activation and that ABC genes are redundantly regulated.

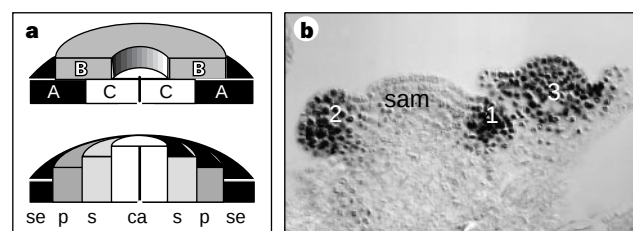
Here we show that the role of *LFY* in flower-meristem initiation can be separated from a role in the later activation of homeotic genes, and that different mechanisms are used in the activation of

three representative members of each of the ABC gene classes. These observations allow us to propose a genetic framework for the patterning of flowers.

## An activated form of *LFY*

In contrast to *LFY* RNA<sup>9</sup>, *LFY* protein persists throughout the flower until floral stage 3, at which time the pattern of ABC genes develop (Fig. 1b). The sequence of *LFY* is not similar to that of other proteins with known biochemical function<sup>9,12</sup>. *LFY* localizes to the nucleus (Fig. 1b), binds DNA in a sequence-specific manner (Fig. 2a), and can, when fused to a heterologous activation domain, mediate transcriptional activation in yeast (Fig. 2b). Although these results indicate that *LFY* is a transcriptional regulator, *in vivo* targets of *LFY* have not been identified so far.

As at least some of its putative downstream genes are expressed in specific patterns within the flower, but *LFY* itself is not, *LFY* activity is likely to be regulated in some way. This regulation could occur at two different levels at least, affecting *LFY*'s DNA-binding or transcriptional-activation potential. To determine whether any of the known ABC genes are likely targets of *LFY*, we generated a version of *LFY* whose transcriptional-activation potential should be constitutive. In this new allele, called *LFY:VP16*, a fusion of *LFY* to the strong activation domain from the viral transcription factor VP16 (refs 13, 14) is expressed under the control of normal *LFY* regulatory sequences<sup>15</sup> (Fig. 2c). The rationale for this experiment was as follows: if *LFY* acts only to specify flower meristem fate, *LFY:VP16* might affect the initial establishment of flower primordia, but not downstream events such as specification of floral organ identity. On the other hand, if *LFY* has a separate role in regulating



**Figure 1** Activity domains of ABC floral homeotic genes and *LFY* expression. **a**, The ABC model<sup>1</sup>. Top, activity domains of ABC genes; bottom, readout of ABC gene activities in form of organ identity. se, sepal; p, petal; s, stamen; ca, carpel. **b**, Immunolocalization of *LFY* protein<sup>26</sup>, which is found in nuclei and expressed fairly uniformly in flower primordia through stage 3 of development. Numbers indicate floral stages. sam, shoot apical meristem.

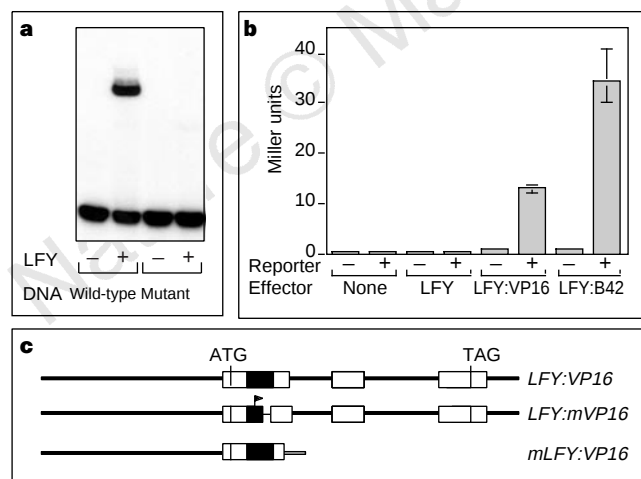
<sup>†</sup> Present addresses: Institut des Sciences Végétales, CNRS, 91198 Gif-sur-Yvette Cedex, France (F.P.); Swedish University of Agricultural Sciences, S-901 83 Umeå, Sweden (O.N.); Department of Biology and Research Center for Cell Differentiation, Seoul National University, Seoul 151-742, Korea (I.L.).



ABC gene expression, then *LFY:VP16* might modify the expression of individual ABC genes and thus affect flower morphology. A caveat to this approach is that, in those cases in which binding of LFY to specific gene promoters is regulated, the VP16 fusion protein might not change expression of target genes.

*LFY:VP16* plants resembled neither *lfy* mutants nor plants that express *LFY* ectopically<sup>8,16</sup>. The primary transformants, presumed to be hemizygous, fell into three phenotypic classes (Fig. 3). Plants with a weak phenotype had some flowers with staminoid petals in the second whorl. Plants with an intermediate phenotype had sepals that were carpeloid, petals that were converted into stamens and reduced in number, and a slight increase in carpel number in the fourth whorl. Finally, in plants with a strong phenotype, each flower was replaced by a carpeloid structure that lacked whorled phyllotaxy. The *LFY:VP16* phenotype was dependent on transgene copy number, such that an intermediate or strong phenotype was seen in homozygous progeny of weak or intermediate plants, respectively. Homozygous progeny of strong transformants could not be recovered because of sterility.

Several controls confirmed that the *LFY:VP16* phenotypes were related to endogenous *LFY* function. First, *LFY:VP16* interacted with wild-type *LFY* in a dosage-dependent manner, as the severity of the *LFY:VP16* phenotype increased when the copy number of endogenous *LFY* was reduced (by crossing *LFY:VP16* to plants with the *lfy-12* null allele). This effect indicates that *LFY:VP16* and wild-type *LFY* compete for the same targets. Furthermore, *LFY:VP16* rescued the flower initiation defects of homozygous *lfy-12* mutants. Finally, two mutant *LFY:VP16* versions (Fig. 2c), in which either the *LFY* coding sequence was truncated downstream to the VP16 domain (*mLFY:VP16*), or the normal VP16 domain was replaced by a mutant derivative shown to be inactive in other systems<sup>17</sup> (*LFY:mVP16*), did not cause any floral phenotype. The



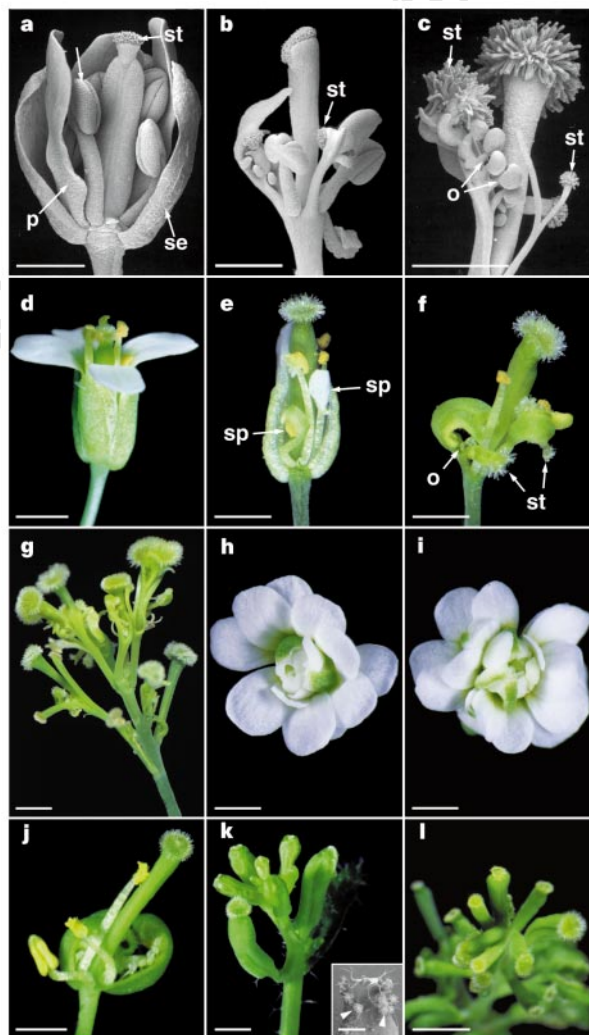
**Figure 2** *In vitro* activity of LFY protein and LFY:VP16 constructs. **a**, Wild-type LFY binds *in vitro* to a double-stranded oligonucleotide (see Methods), as seen by a shift in electrophoretic mobility on a non-denaturing polyacrylamide gel. A two-base-pair change in the DNA sequence (mutant) abolishes DNA binding. The binding site was originally identified by immunoprecipitation of genomic DNA from the *AP1* locus. **b**, LFY activates gene expression in the yeast *Saccharomyces cerevisiae*. LFY was expressed together with a *lacZ* reporter, which was controlled by a yeast minimal promoter linked to a 194 bp DNA fragment spanning the LFY-binding site shown in **a**. The reporter is activated if LFY is fused to a strong transcriptional activation domain, such as the one from VP16 (ref. 13) or the synthetic domain B42 (ref. 44). Bars indicate the range of values from three replicate transformations. **c**, Diagram of *LFY:VP16* transgenes. Exons are indicated by rectangles. Sequences coding for the VP16 activation domain are shown in black. The flag in *LFY:mVP16* indicates a missense mutation in the mutant VP16 domain<sup>17</sup>.

*LFY:mVP16* transgene complemented all floral defects of the *lfy-12* null allele, showing that insertion of a foreign protein domain did not inactivate the LFY protein.

### Differential effects of LFY:VP16 on ABC genes

As the homeotic transformations in *LFY:VP16* flowers indicated a change in ABC gene expression, we analysed RNA expression of representatives of each class of homeotic genes.

The A-function gene *AP1* is expressed uniformly in early wild-type flower buds, and later becomes restricted to the two outer whorls because of repression by *AG*<sup>4,18</sup>. In *LFY:VP16* flowers, the early pattern of *AP1* was unchanged, but the expression level was



**Figure 3** Phenotypes of wild-type, mutant and transgenic plants. **a**, Scanning electron micrograph (SEM) of a wild-type flower, with one sepal, petal and stamen removed. **b**, SEM of a flower with an intermediate *LFY:VP16* phenotype. Note the carpeloid of first-whorl organ. **c**, SEM of a flower with a strong *LFY:VP16* phenotype. Most organs are tipped with stigmatic tissue. Ovules are visible on one organ. **d**, Light micrograph of a wild-type flower. **e**, Flower with a weak *LFY:VP16* phenotype, with one sepal removed. Second-whorl petals show partial conversion into stamens. **f**, Flower with an intermediate *LFY:VP16* phenotype. Some of the first-whorl sepals are carpeloid, indicated by ovules and stigmatic tissue. **g**, Inflorescence with a strong *LFY:VP16* phenotype. **h**, *ag-1* flower. **i**, *LFY:VP16 ag-1* flower. **j**, *ap2-2* flower. **k**, *LFY:VP16 ap2-2* inflorescence. The inset shows a cauline leaf with stigmatic tissue on its margins (arrow heads). **l**, *35S::LFY ap2-2* inflorescence. se, sepal; p, petal; s, stamen; st, stigmatic tissue; sp, staminoid petal; o, ovule. Scale bars, 500  $\mu$ m in **a-c** and inset in **k**, and 1 mm in **d-l**.

greatly increased (Fig. 4a, e).

The C-function gene *AG* begins to be expressed in the centre of wild-type flowers during stage 3 (ref. 19). Both the spatial and the temporal patterns of *AG* RNA were changed markedly in *LFY:VP16* flowers. *AG* expression began earlier and was detected throughout the flower (Fig. 4b, f). Comparison of *AP1* and *AG* expression patterns indicates that *LFY:VP16* may be able to override repression of *AP1* by *AG*<sup>4</sup>, but probing of adjacent sections for *AP1* and *AG* expression will be needed to confirm this. In addition to changes in the pattern of expression of *AG*, the levels of *AG* expression were increased in *LFY:VP16* plants.

As the conversion of sepals into carpels and of petals into stamens in *LFY:VP16* plants was similar to the phenotype of plants that express *AG* constitutively<sup>20</sup>, we crossed plants with the strong *ag-1* allele to *LFY:VP16* plants. Sepal and petal identity were restored in *LFY:VP16 ag-1* plants, showing that ectopic *AG* expression was the main cause of the *LFY:VP16* phenotype (Fig. 3h, i).

The B-function gene *AP3* is first expressed in wild-type flowers during stage 3, and its expression is largely restricted to the presumptive second and third whorls<sup>21</sup>. *LFY:VP16* had only minor effects on *AP3* expression. *AP3* expression was normal in the lateral regions of flowers, but reduced in medial regions, correlating with missing second- and third-whorl organs in these positions (Fig. 4c, g). *AP3* expression in first-whorl organs of advanced *LFY:VP16* flowers was also more extensive than in wild-type flowers (results not shown).

Thus an activated version of *LFY*, *LFY:VP16*, has no pronounced effect on one ABC gene, *AP3*, but causes increased activation of another ABC gene, *AP1*, and both increased and ectopic activation of a third ABC gene, *AG*. If *LFY* controlled homeotic gene expression only through its role in specifying flower meristem fate, we would expect that changing the transcriptional-activation potential of *LFY* might affect flower initiation but not the later activation of ABC genes. As an alteration in the activation potential of *LFY* affects ABC gene expression, we propose that *LFY* has a role in activating ABC genes that is separable from its role in specifying flower meristem identity. Furthermore, the differential effects of *LFY:VP16* on different ABC genes indicate that *LFY* may interact with these genes through different mechanisms.

### Flower-independent activation of *AP1*

*LFY* is expressed in both leaf and flower primordia, at lower levels in the former<sup>15</sup>. As ABC genes are not normally expressed in leaves, we tested whether this was due to insufficient levels of *LFY*, or to the

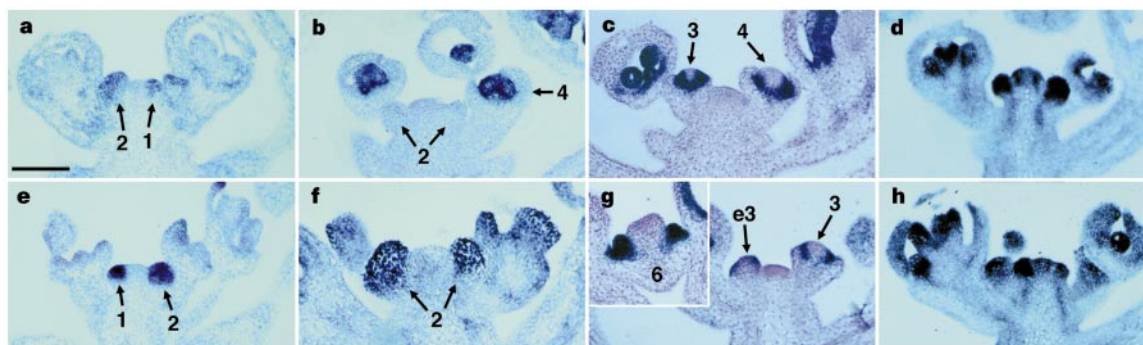
absence of other, flower-specific factors, whose activity could possibly be mimicked by adding the VP16 domain to *LFY*. We introduced a transgene that causes *LFY* overexpression in vegetative tissues (35S::*LFY*; ref. 16) into an *AP1*::*GUS* reporter line, and found that the *AP1* promoter is induced in young 35S::*LFY* seedlings before there is any sign of ectopic flower formation (Fig. 5a–c). We verified that endogenous *AP1* RNA was induced in young 35S::*LFY* seedlings by reverse transcription followed by polymerase chain reaction (RT-PCR) (Fig. 5c). Thus the role of *LFY* in inducing *AP1* can be separated from its role in the specification of flower meristems.

Morphological studies have shown that, after floral induction in wild-type *Arabidopsis*, primordia that would otherwise have become leaves develop into flowers instead<sup>22</sup>. Interestingly, *AP1* promoter activity in 35S::*LFY* seedlings was mainly confined to young leaf primordia (Fig. 5c), indicating that *LFY* may interact with other differentially active factors to induce *AP1* expression preferentially in primordia that have the potential to adopt a floral fate.

### Flower-independent activation of *AG*

In contrast to *AP1*, *AG* is not induced in vegetative tissue of 35S::*LFY* plants, as shown by *in situ* hybridization (results not shown) and with an *AG*::*GUS* reporter line (Fig. 5e), indicating that *AG* activation does not depend on *LFY* levels only. However, the ability of *LFY:VP16* to activate *AG* throughout the flower meristem indicated that the VP16 domain may allow *LFY* to function largely independently of other factors that regulate *AG*. If this were correct, *LFY:VP16* should be able to induce *AG* expression in vegetative tissue. Like endogenous *LFY*, *LFY:VP16* should be expressed, albeit at low levels, in leaf primordia, especially in those giving rise to cauline leaves, which are initiated just before the first flower is formed<sup>15</sup>. Consistent with a weak effect of *LFY:VP16* on *AG* in cauline leaves, we occasionally observed stigmatic tissue, normally found only at the tip of carpels, along the margin of cauline leaves (results not shown). The appearance of stigmatic tissue in leaves was much more pronounced when *LFY:VP16* was introduced into *ap2* mutants (Fig. 3k). *AP2* is functionally a flower-specific repressor of *AG*, although *AP2* is expressed throughout the plant<sup>23</sup>. The enhancement of the *LFY:VP16* leaf phenotype in *ap2* mutants indicates that low levels of *LFY:VP16* present in leaves do not normally overcome repression of *AG* by *AP2*.

To test more directly whether *LFY:VP16* can induce *AG* in any vegetative tissue, we tried to generate plants in which *LFY:VP16* was



**Figure 4** Expression of *AP1*, *AG*, *AP3* and *LFY* in wild-type (a–d) and *LFY:VP16* transgenic (e–h) plants. Longitudinal sections of flowering apices that were hybridized *in situ* are shown. *LFY:VP16* sections were from plants with intermediate (e, g, h) or strong (f) phenotypes. a, e, *AP1* expression during stages 1 and 2 is increased in *LFY:VP16* plants. b, f, *AG* RNA is first expressed in the centre of wild-type flowers during stage 3. In *LFY:VP16* plants, *AG* expression is detected during stages 1 and 2 and extends into the subapical region of the shoot meristem, reminiscent of the *LFY* expression pattern (d, h). *AG* and *AP1*

RNAs (e) seem to overlap; however, the sections probed for *AP1* and *AG* were from *LFY:VP16* lines of different phenotypic strengths. c, g, *AP3* expression, which in wild-type plants begins during early stage 3 (e3), is restricted to whorls 2 and 3 and the base of whorl 1. In *LFY:VP16* flowers, *AP3* expression is unchanged in the lateral part of the meristem (inset) and slightly reduced in the medial part (main panel). d, h, *LFY* expression is similar in wild-type and *LFY:VP16* inflorescences. Scale bar, 100  $\mu$ m.

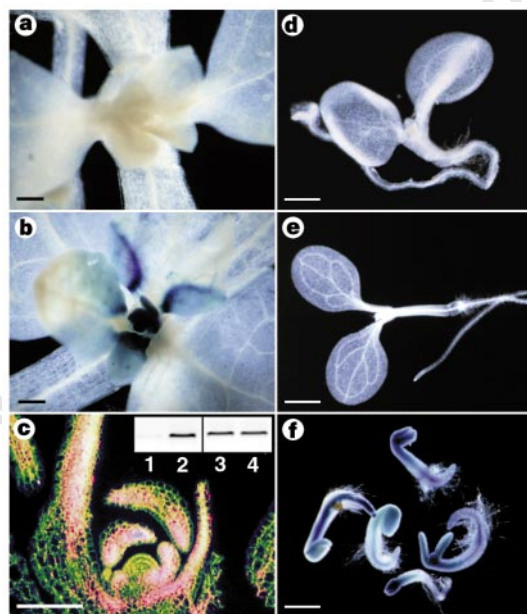


constitutively expressed (*35S::LFY:VP16*). We never recovered adult transformants, so we vacuum-infiltrated<sup>24</sup> an *Agrobacterium* strain containing the *35S::LFY:VP16* vector directly into homozygous *AG::GUS* reporter plants. About 30,000 seeds were collected, germinated and stained for GUS activity. Although no GUS activity was detected in 20,000 control *AG::GUS* seedlings, about 30 seedlings derived from vacuum infiltration with *35S::LFY:VP16* were growth-arrested and showed GUS activity in all organs (Fig. 5f). This frequency, 0.1%, concurs with our average transformation efficiencies of between 0.05% and 2%. We conclude that, once strongly expressed, *LFY:VP16* can induce *AG* independently of flower initiation.

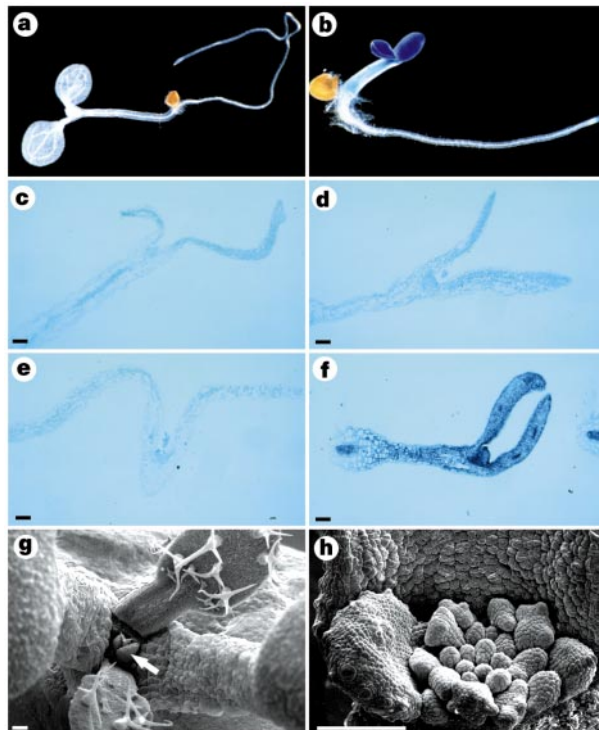
The observation that the *VP16* activation domain enables *LFY* to induce *AG* ectopically, both in vegetative tissues and in flowers, does not indicate whether *LFY* is normally an activator or a repressor of *AG*. It has been suggested that *LFY* may activate *AG* during early flower development but contributes subsequently to repression of *AG* in the outer whorls<sup>6,7,12</sup>. We therefore re-examined *35S::LFY* plants, in which *LFY* levels are increased throughout the flower<sup>16</sup>. Although most *35S::LFY* flowers were phenotypically wild type, the most apical ones produced in short days had carpeloid organs in the first whorl and lacked some petals, indicating increased *AG* expression. This was even more obvious in plants resulting from the cross of *35S::LFY* to *ap2-2* and *ap2-1* mutants, in which a synergistic effect was seen (Fig. 3j; results not shown). *35S::LFY ap2-2* flowers resembled those of *LFY:VP16 ap2-2* plants (Fig. 3k, l). The fact that increased carpeloidy is caused by *LFY* overexpression in *ap2* mutants is complementary to the observation that eliminating *LFY* function

in *lfy ap2* double mutants reduces carpeloidy as compared with *ap2* single mutants<sup>9,10</sup>. It therefore seems that the primary role of *LFY* is activation of *AG*.

Why does *LFY*, which is present throughout the flower, normally induce *AG* only in the centre of the flower? In the prevailing view of *AG* regulation, meristem-identity genes such as *LFY* activate *AG* throughout the flower, but *AG* is specifically repressed in the outer whorls by A-function genes such as *AP2* (refs 1, 7). An alternative, formally equivalent model of *AG* regulation holds that A-function genes are general repressors of *AG*, and that this repression is selectively overcome in the centre of the flower during stage 3. We propose that a combination of the two models is most likely to reflect the *in vivo* situation. For example, if *AG* were uniformly activated throughout the flower, and if absence of *AG* RNA from outer whorls was caused by repression by *AP2*, then *AG* should be uniformly expressed in *ap2* mutant flowers. However, *AG* expression is weaker in the first whorl of *ap2* mutants than in interior whorls<sup>19</sup>. Furthermore, although phenotypic analysis indicates that *AP2* functions throughout the flower in *ag* mutants, *AP2* cannot repress *AG* RNA expression in the centre of these flowers<sup>1,4</sup>. These results indicate that *AG* may be induced to a greater extent in the centre of the flower than in the periphery. We propose that this differential effect involves the interaction of *LFY* with a co-activator that aids *LFY* in overcoming *AP2*-mediated repression of *AG* in the centre of the flower. The *VP16* activation domain may render *LFY* independent of other regulators of *AG*, thus allowing *LFY* to induce *AG* throughout the flower even in the presence of repressors such as *AP2*.



**Figure 5** *AP1::GUS* and *AG::GUS* expression outside flowers. **a–c**, *AP1::GUS* activity in 11-day-old plants. **a**, Wild-type plants have no detectable *AP1::GUS* activity. **b**, *35S::LFY* plants express *AP1::GUS* throughout young leaf primordia and at the base of older leaves. **c**, Longitudinal section through the shoot apex of a *35S::LFY* plant. GUS staining, which appears pink under dark field microscopy, is present at high levels in leaves. It is either absent from the shoot apical meristem, or present at much lower levels. (Meristematic cells in the centre are not vacuolated and appear yellow even in control plants.) The inset shows detection of RNA by RT-PCR in young seedlings (six days after germination, grown in short days). Lanes 1, 2, *AP1* in wild-type and *35S::LFY* plants, respectively; lanes 3, 4, control amplification of *eIF4A* in wild-type and *35S::LFY* plants, respectively. **d–f**, *AG::GUS* activity is absent from 5-day-old wild-type (**d**) and *35S::LFY* (**e**) plants, but is detected in all organs of *35S::LFY:VP16* seedlings, which are growth-arrested (**f**). Scale bars, 200  $\mu$ m in **a**, **b**, 100  $\mu$ m in **c**, and 500  $\mu$ m in **d–f**.



**Figure 6** *LFY* and *UFO* together induce *AP3* expression in seedlings. **a**, **b**, *AP3::GUS* is expressed in *35S::LFY 35S::UFO* seedlings (**b**), but not in wild-type seedlings (**a**). **c–f**, *In situ* hybridization does not detect *AP3* RNA in 5-day-old seedlings with no transgene (**c**), only *35S::LFY* (**d**), or only *35S::UFO* (**e**). In contrast, *AP3* RNA is detected in *35S::LFY 35S::UFO* seedlings (**f**). **g**, SEM of a shoot apex of a 5-day-old non-transgenic seedling. Leaf primordia cover the shoot meristem (arrow). **h**, SEM of a shoot apex of a *35S::LFY 35S::UFO* seedling. The shoot meristem has been consumed by primordia whose epidermal cells are reminiscent of petals or stamens. Scale bars, 1 mm in **a**, **b**, and 100  $\mu$ m in **c–h**.



# Flower-independent activation of AP3

Activation of the A-function gene *API* or of the C-function gene *AG* can be separated from flower initiation by overexpression of normal *LFY* or of an activated form of *LFY*. However, in neither of these situations did we observe a marked change in activation of the B-function gene *AP3*, indicating that regulation of *AP3* by *LFY* relies on a mechanism different from the ones used for *API* and *AG*. A partially redundant co-regulator that might act together with *LFY* in activating *AP3* is *UNUSUAL FLORAL ORGANS (UFO)*. When expressed constitutively, *UFO* causes ectopic activation of *AP3* within flowers<sup>25</sup>. All *UFO* loss- and gain-of-function phenotypes are masked in a *lfy* null background<sup>25–27</sup>, indicating that *UFO* activity requires functional *LFY* protein. Furthermore, *LFY* and *UFO* expression domains overlap at the time of *AP3* induction, indicating that *LFY* and *UFO* protein may act together, with *UFO* providing much of the positional information for *AP3* activation. The molecular function of *UFO* and of its snapdragon orthologue *FIMBRIATA* is unknown, but the presence of a functional F-box indicates that they may be involved in targeting other proteins for ubiquitin-mediated degradation<sup>28–30</sup>. Thus, although it is unlikely that *UFO* itself is a transcriptional regulator, like some other F-box proteins *UFO* might regulate the activity of transcription factors<sup>31,32</sup>.

To determine whether *LFY* and *UFO* together can induce *AP3* independently of flower initiation, we generated doubly transgenic seedlings that constitutively expressed both *LFY* and *UFO*. *35S::LFY 35S::UFO* seedlings were growth-arrested and did not form any mature leaves. The combination of *35S::LFY* and *35S::UFO* transgenes was sufficient to induce *AP3* in vegetative tissues of seedlings (Fig. 6). Thus activation of *AP3*, like that of *API* and *AG*, can be separated from flower initiation.

## A genetic framework for floral patterning

We have shown, for representatives of each of the three ABC classes of floral homeotic genes, that their activation by *LFY* can be uncoupled from the initiation of flowers, and we conclude that *LFY* not only confers flower-meristem identity at an early stage of flower development, but also has a different role in the later activation of ABC genes. Because of the early role of *LFY* in flower development, and because of redundancy of *LFY* with other genes, we used untraditional approaches, such as using transgenic plants that express an activated version of *LFY*, to dissect the different roles of *LFY*. This strategy, applied previously to transcription factors from *Caenorhabditis elegans*<sup>33</sup> and *Drosophila melanogaster*<sup>34</sup>, has proven very informative, and could be used in the study of many plant transcription factors. However, our gain-of-function studies were guided by previous loss-of-function analyses, and loss-of-function data together with careful controls are essential to validate results generated with the kind of approaches used here. Interestingly, the VP16 fusion did not equally affect all known

functions of *LFY*, including flower initiation or activation of B-function genes. This could be because the binding of *LFY* to target sites in certain promoters is regulated, or because a minimal concentration of *LFY* is required for cooperative DNA binding to selected sites, in which case adding the VP16 domain would have no effect below a threshold of *LFY*.

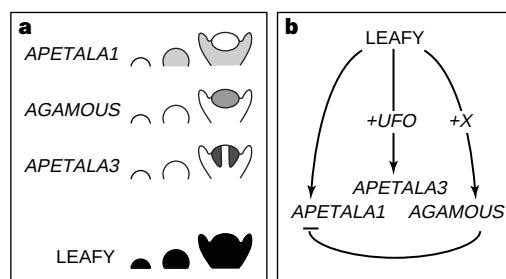
Our results indicate that different mechanisms may regulate each of the three ABC genes that we have identified as *LFY* targets (Fig. 7). The A-function gene *API* is initially expressed uniformly throughout floral primordia, where it is activated shortly after *LFY* activation<sup>9,18</sup>. As *LFY* can induce *API* very early during the formation of any primordium that could potentially become a flower, no region- or flower-specific co-regulator is needed for *API*. In contrast to *API*, both *AG* and *AP3* are activated in region-specific patterns within flowers<sup>19,21</sup>. Activation of *AP3* seems to rely mainly on the combination of *LFY*, which is uniformly expressed in flower meristems, and other factors such as *UFO*, which is expressed in a region-specific pattern in both shoot and flower meristems. In this case, *LFY* provides flower-meristem specificity and *UFO* provides region specificity. A similar situation could apply to *AG*, that is, *LFY* may act in combination with another factor that is expressed in the centre of both shoot and flower meristems. Note that shoot meristems resemble stage 2 flower meristems, both morphologically and at the level of gene expression. In addition to *UFO*, several other genes are expressed in similar patterns in shoot and stage 2 flower meristems, although mutations in these genes—in contrast to mutations in *UFO*—disrupt development of both shoots and flowers<sup>35–37</sup>.

A scenario that emerges from these observations is that the patterning of ABC gene expression is achieved by co-opting a meristem patterning system that evolved before flowers appeared. To this underlying pattern, which is common to shoot and flower meristems, are added a few flower-specific factors that are not region-specific themselves, such as *LFY*, and the combination of these two classes of factors provides region specificity within flowers. This model may explain why the search for specific regulators of ABC gene expression has met with limited success. If most factors controlling region-specific expression of ABC genes also affect shoot development, it might be difficult to uncover them in conventional genetic screens. Other unresolved problems include the molecular mechanisms by which *LFY* controls the expression of target genes; the mechanistic basis of redundancy of *LFY* and *UFO* with other meristem-identity and patterning genes; and the nature of the interaction of *LFY* with other important floral regulators, such as *AP2*, *LEUNIG* or *CURLY LEAF*<sup>23,38,39</sup>. □

## Methods

**Plasmid constructs.** Plant expression: *LFY:VP16* was generated by inserting a segment encoding the VP16 activation domain (amino acids 413–490), amplified by PCR from plasmid pRG50 (ref. 14), into a *LFY* genomic fragment (nucleotides 465–2,844, 2,845–5,937 of GenBank accession number M91208 (ref. 9)). The *LFY:VP16* chimera was cloned into the transformation vector pCGN1547 (ref. 40), producing pDW245. *LFY:mVP16* (pFP21) was generated in the same way, using the sequence encoding a mutant, truncated VP16 activation domain (amino acids 413–456), originating from plasmid pMSVP16 ΔC119/FP442 (ref. 17). To generate *mLFY:VP16* (pFP17), we used PCR to replace the region 3' to *VP16* sequences in pDW245 by a stop codon followed by the transcriptional terminator of the nopaline synthase gene. For *35S::LFY:VP16*, a chimera containing *LFY* complementary DNA and the wild-type *VP16* activation domain was generated in pBluescriptKS<sup>+</sup> (pFP10). The *LFY:VP16* cDNA was then inserted into the 35S vector pCGN18 (ref. 41), to yield pFP15.

**Bacterial expression:** The *LFY* open-reading frame, encoding a variant with a deletion of unconserved amino acids 391–421, was cloned into the *Escherichia coli* expression vector pET28a<sup>+</sup> (Novagen), to yield pMX013. The deletion has minimal effects on *LFY* activity, as assayed by constitutive expression in plants (pIL12), but allows *LFY* expression in bacteria.



**Figure 7** Model for ABC gene activation by *LFY*. **a**, Comparison of the expression profiles of *LFY* protein and ABC gene RNAs. Left, stage 0 flower; middle, stage 1 and 2 flower; right, stage 3 flower. **b**, Effect of *LFY* on different targets. The VP16 activation domain makes *LFY* independent of factor 'X'.

**Yeast expression:** We cloned a 194-base-pair (bp) *BglIII/SalI* fragment from the *API* promoter into pLG178 (ref. 42), to generate the reporter pFP30; *LFY* and *LFY:VP16* cDNAs were subcloned into p424 (ref. 43), to generate the *LFY* and *LFY:VP16* effector constructs (pFP13 and pFP14, respectively). We cloned a *LFY* cDNA into pJG4-5 (ref. 44), to generate the *LFY:B42* fusion (pDW188).

**In vitro studies.** *LFY* protein with amino- and carboxy-terminal hexahistidine tags was expressed in *E. coli* BL21-DE3 cells. Insoluble protein was isolated as inclusion bodies, denatured, affinity-purified over a Ni-NTA (Qiagen) column, and renatured by stepwise dialysis against decreasing concentrations of urea. Oligonucleotides were labelled with T4-poly-nucleotide kinase and [ $\gamma$ - $^{32}$ P]ATP. The binding reaction, with 25 fmol of target DNA, was done for 10 min on ice in 20 mM Tris pH 7.5, 150 mM NaCl, 0.25 mM EDTA, 20% glycerol, 1 mM dithiothreitol, 20 mM MgCl<sub>2</sub>, 0.02% NP-40 and 12.5 ng  $\mu$ l<sup>-1</sup> poly(dI-dC). The wild-type sequence for the electrophoretic mobility shift assay was 5'-TTG GGG AAG GAC CAG TGG TCC GTA CAA TGT-3', the italicized bases were changed to AA in the mutant. Yeast transformations and  $\beta$ -galactosidase activity measurements were done as described<sup>44,45</sup>.

**Plant material.** The following transgenic lines were used: DW151.2.5L and DW151.2.5C (35S::LFY) (refs 16, 46); DW229.5.3 (35S::UFO) (ref. 25); AM154.5c (*API*::GUS) (ref. 47); 1008.5 (*AP3*::GUS) (ref. 48); and pAG-I::GUS (*AG*::GUS) (ref. 49). GUS assays<sup>15</sup> were done with material grown in short days.

**Expression analyses.** We generated antisense RNA probes for *in situ* hybridization from plasmids pDW119 (*LFY*) (ref. 9), pCIT565 (*AG*) (ref. 19), pD793 (*AP3*) (ref. 21), and pAM128 (*API*) (ref. 18). Hybridization and signal detection were as described<sup>50</sup> (see also <http://www.wisc.edu/genetics/CATG/barton/protocols.html>), except that RNase treatment was omitted and counter-staining of some sections was done with basic fuchsin. For RT-PCR, RNA was extracted from 6-day-old seedlings grown in short days using the RNeasy Plant Mini Kit (Qiagen). RT-PCR was carried out with the Titan RT-PCR kit (Boehringer). An *API* cDNA fragment was detected using oligonucleotides 5'-GCA CAT CCG CAC TAG AAA AAA CCA AC-3' and 5'-CTT CTT GAT ACA GAC CAC CCA TGT-3'. As a control, a fragment from the gene encoding eIF4A was amplified using 5'-TTC TCA AAC CAT AAG CAT AAA TAC CC-3' and 5'-AAA CTC AAG GAA GTA CTT GAG GGA CAA-3'. Aliquots were analysed after 25 cycles.

**Scanning electron microscopy.** Tissue was prepared as described<sup>1</sup>, and viewed in a Cambridge S360 microscope at an accelerating voltage of 10 kV.

Received 27 April; accepted 16 July 1998.

- Bowman, J. L., Smyth, D. R. & Meyerowitz, E. M. Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112**, 1–20 (1991).
- Coen, E. S. & Meyerowitz, E. M. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**, 31–37 (1991).
- Weigel, D. & Meyerowitz, E. M. The ABCs of floral homeotic genes. *Cell* **78**, 203–209 (1994).
- Gustafson-Brown, C., Savidge, B. & Yanofsky, M. F. Regulation of the *Arabidopsis* floral homeotic gene *APETALA1*. *Cell* **76**, 131–143 (1994).
- Yanofsky, M. Floral meristems to floral organs: genes controlling early events in *Arabidopsis* flower development. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **46**, 167–188 (1995).
- Hantke, S. S., Carpenter, R. & Coen, E. S. Expression of *floricaula* in single cell layers of periclinal chimeras activates downstream homeotic genes in all layers of floral meristems. *Development* **121**, 27–35 (1995).
- Weigel, D. & Meyerowitz, E. M. Activation of floral homeotic genes in *Arabidopsis*. *Science* **261**, 1723–1726 (1993).
- Schultz, E. A. & Haughn, G. W. *LEAFY*, a homeotic gene that regulates inflorescence development in *Arabidopsis*. *Plant Cell* **3**, 771–781 (1991).
- Weigel, D., Alvarez, J., Smyth, D. R., Yanofsky, M. F. & Meyerowitz, E. M. *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell* **69**, 843–859 (1992).
- Huala, E. & Sussex, I. M. *LEAFY* interacts with floral homeotic genes to regulate *Arabidopsis* floral development. *Plant Cell* **4**, 901–913 (1992).
- Gustafson-Brown, C. *Characterization of the Arabidopsis Floral Homeotic Gene APETALA1*. Thesis, Univ. California, San Diego (1996).
- Coen, E. S. *et al.* *floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell* **63**, 1311–1322 (1990).
- Trizezenberg, S. J., Kingsbury, R. C. & McKnight, S. L. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev.* **2**, 718–729 (1988).
- Cousens, D. J., Greaves, R., Goding, C. R. & O'Hare, P. The C-terminal 79 amino acids of the herpes simplex virus regulatory protein,  $\gamma$ mw65, efficiently activate transcription in yeast and mammalian cells in chimeric DNA-binding proteins. *EMBO J.* **8**, 2337–2342 (1989).
- Blázquez, M. A., Soowal, L., Lee, I. & Weigel, D. *LEAFY* expression and flower initiation in *Arabidopsis*. *Development* **124**, 3835–3844 (1997).

- Weigel, D. & Nilsson, O. A developmental switch sufficient for flower initiation in diverse plants. *Nature* **377**, 495–500 (1995).
- Cress, W. D. & Trizezenberg, S. J. Critical structural elements of the VP16 transcriptional activation domain. *Science* **251**, 87–90 (1991).
- Mandel, M. A., Gustafson-Brown, C., Savidge, B. & Yanofsky, M. F. Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* **360**, 273–277 (1992).
- Drews, G. N., Bowman, J. L. & Meyerowitz, E. M. Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the *APETALA2* product. *Cell* **65**, 991–1002 (1991).
- Mizukami, Y. & Ma, H. Ectopic expression of the floral homeotic gene *AGAMOUS* in transgenic *Arabidopsis* plants alters floral organ identity. *Cell* **71**, 119–131 (1992).
- Jack, T., Brockman, L. L. & Meyerowitz, E. M. The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS-box and is expressed in petals and stamens. *Cell* **68**, 683–697 (1992).
- Hempel, F. D. & Feldman, L. J. Bi-directional influence development in *Arabidopsis thaliana*: acropetal initiation of flowers and basipetal initiation of paracoles. *Planta* **192**, 276–286 (1994).
- Jofuku, K. D., den Boer, B. G. W., Van Montagu, M. & Okamura, J. K. Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. *Plant Cell* **6**, 1211–1225 (1994).
- Bechtold, N., Ellis, J. & Pelletier, G. In planta *Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *C.R. Acad. Sci.* **316**, 1194–1199 (1993).
- Lee, I., Wolfe, D. S., Nilsson, O. & Weigel, D. A *LEAFY* co-regulator encoded by *UNUSUAL FLORAL ORGANS*. *Curr. Biol.* **7**, 95–104 (1997).
- Levin, J. Z. & Meyerowitz, E. M. *UFO*: an *Arabidopsis* gene involved in both floral meristem and floral organ development. *Plant Cell* **7**, 529–548 (1995).
- Wilkinson, M. D. & Haughn, G. W. *UNUSUAL FLORAL ORGANS* controls meristem identity and organ primordia fate in *Arabidopsis*. *Plant Cell* **7**, 1485–1499 (1995).
- Feldman, R. M. R., Correll, C. C., Kaplan, K. B. & Deshaies, R. J. A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* **91**, 221–230 (1997).
- Skowra, D., Craig, K. L., Tyers, M., Elledge, S. J. & Harper, J. W. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91**, 202–219 (1997).
- Ingram, G. C. *et al.* Dual role for fimbriata in regulating floral homeotic genes and cell division in *Antirrhinum*. *EMBO J.* **16**, 6521–6534 (1997).
- Kumar, A. & Paietta, J. V. An additional role for the F-box motif: gene regulation within the *Neurospora crassa* sulfur control network. *Proc. Natl. Acad. Sci. USA* **95**, 2417–2422 (1998).
- Patton, E. E. *et al.* Cdc53 is a scaffold protein for multiple Cdc34/Skp1/F-box protein complexes that regulate cell division and methionine biosynthesis in yeast. *Genes Dev.* **12**, 692–705 (1998).
- Sze, J. Y., Liu, Y. & Ruvkun, G. VP16-activation of the *C. elegans* neural specification factor UNC-86 suppresses mutations in downstream genes and causes defects in neural migration and axon outgrowth. *Development* **124**, 1159–1168 (1997).
- Jimenez, G., Pinchin, S. M. & Ish-Horowitz, D. *In vivo* interactions of the *Drosophila* Hair and Runt transcriptional repressors with target promoters. *EMBO J.* **15**, 7088–7098 (1996).
- Long, J. A., Moan, E. L., Medford, J. I. & Barton, M. K. A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene in *Arabidopsis*. *Nature* **379**, 66–69 (1996).
- Souer, E., van Houwelingen, A., Kloos, D., Mol, J. & Koes, R. The *no apical meristem* gene of petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell* **85**, 159–170 (1996).
- Clark, S. E., Williams, R. W. & Meyerowitz, E. M. The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* **89**, 575–585 (1997).
- Liu, Z. & Meyerowitz, E. M. *LEUNIG* regulates *AGAMOUS* expression in *Arabidopsis* flowers. *Development* **121**, 975–991 (1995).
- Goodrich, J. *et al.* A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature* **386**, 44–48 (1997).
- McBride, K. E. & Summerfelt, K. R. Improved binary vectors for *Agrobacterium*-mediated plant transformation. *Plant Mol. Biol.* **14**, 269–276 (1990).
- Jack, T., Fox, G. L. & Meyerowitz, E. M. *Arabidopsis* homeotic gene *APETALA3* ectopic expression: transcriptional and posttranscriptional regulation determine organ identity. *Cell* **76**, 703–716 (1994).
- Guarente, L. & Mason, T. Heme regulates transcription of the *CYC1* gene of *S. cerevisiae* via an upstream activation site. *Cell* **32**, 1279–1286 (1983).
- Mumberg, D., Muller, R. & Funk, M. Regulatable promoters of *Saccharomyces cerevisiae*; comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.* **22**, 5767–5768 (1994).
- Golemis, E. A., Gyuris, J. & Brent, R. In *Current Protocols in Molecular Biology* (ed. Ausubel, F. M.) 20.1.1–20.1.28 (John Wiley, New York, 1996).
- Gietz, R. D., Schiestl, R. H., Willems, A. R. & Woods, R. A. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* **11**, 355–360 (1995).
- Nilsson, O., Lee, I., Blázquez, M. A. & Weigel, D. Flowering-time genes modulate the response to *LEAFY* activity. *Genetics* (in the press).
- Hempel, F. D. *et al.* Floral determination and expression of floral regulatory genes in *Arabidopsis*. *Development* **124**, 3845–3853 (1997).
- Krizek, B. A. & Meyerowitz, E. M. The *Arabidopsis* homeotic genes *APETALA3* and *PISTILLATA* are sufficient to provide the B class organ identity function. *Development* **122**, 11–22 (1996).
- Sieburth, L. E. & Meyerowitz, E. M. Molecular dissection of the *AGAMOUS* control region shows that *cis* elements for spatial regulation are located intragenically. *Plant Cell* **9**, 355–365 (1997).
- Jackson, D., Veit, D. & Hake, S. Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetable shoot. *Development* **120**, 405–413 (1994).

**Acknowledgements.** We thank D. S. Wolfe, K. Bombli, V. Hedquist and T. Nguyen for technical assistance; M. Blázquez, E. Bier, A. Colon, C. Fankhauser, G. Gocal, M. Goulding, D. Jackson, I. Kardailsky, M. Ng and M. Yanofsky for discussions, technical advice and unpublished results; T. Jack, B. Krizek, A. Mandel and L. Sieburth for reporter lines; P. O'Hare and S. Trizezenberg for VP16 plasmids; and G. Ruvkun for inspiring the VP16 experiment. This work was supported by fellowships from Bourse Lavoisier du Ministère Français des Affaires Étrangères (E.P.), the Human Frontier Science Program Organization (E.P., O.N., M.A.B.), and the Parson and Aron Foundation (I.L.), and by grants from the National Science Foundation and US Department of Agriculture (to D.W.). D.W. is an NSF Young Investigator and receives support from the Samuel Roberts Noble Foundation, Agritope and ForBio.

Correspondence and requests for materials should be addressed to D.W. (e-mail: weigel@salk.edu).

# Crystal structure of a hepatitis delta virus ribozyme

Adrian R. Ferré-D'Amaré\*, Kaihong Zhou† & Jennifer A. Doudna\*†

\* Department of Molecular Biophysics and Biochemistry, and † Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06520-8114, USA

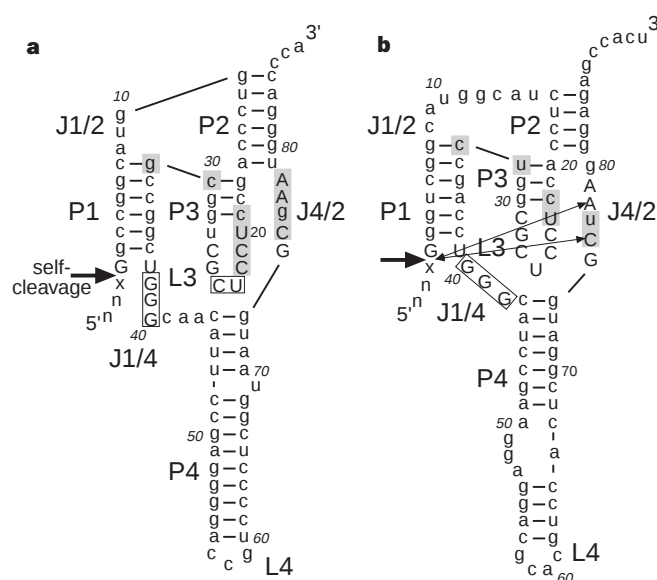
**The self-cleaving ribozyme of the hepatitis delta virus (HDV) is the only catalytic RNA known to be required for the viability of a human pathogen. We obtained crystals of a 72-nucleotide, self-cleaved form of the genomic HDV ribozyme that diffract X-rays to 2.3 Å resolution by engineering the RNA to bind a small, basic protein without affecting ribozyme activity. The co-crystal structure shows that the compact catalytic core comprises five helical segments connected as an intricate nested double pseudoknot. The 5'-hydroxyl leaving group resulting from the self-scission reaction is buried deep within an active-site cleft produced by juxtaposition of the helices and five strand-crossovers, and is surrounded by biochemically important backbone and base functional groups in a manner reminiscent of protein enzymes.**

HDV is an RNA satellite virus of hepatitis B virus (HBV). Infection of humans by both HBV and HDV is generally associated with more severe hepatitis than that caused by HBV alone<sup>1</sup>. HDV has a circular, single-stranded RNA genome of 1,700 nucleotides, with ~70% self-complementarity. This genome is thought to be replicated by the host RNA polymerase II through a double-rolling-circle mechanism<sup>1,2</sup>. Replication produces linear multimers of both genomic and antigenomic RNAs, both of which possess a self-cleaving activity that processes the RNAs to unit length<sup>3–5</sup>. The activity has been mapped to a catalytic-RNA, or ribozyme, domain whose function is required for HDV replication *in vivo*<sup>1</sup>. The HDV ribozyme, which is active *in vitro* in the absence of any proteins, is the only known example of a catalytic RNA associated with an animal virus. The circular single-stranded RNA genome, its mode of replication by a host RNA polymerase, and the self-cleaving-RNA activity of HDV have close parallels in the plant viroids. Apart from the genomic and antigenomic ribozymes, however, there are no known homologues of HDV ribozymes, and sequence variation of the HDV ribozymes in clinical isolates is minimal<sup>2</sup>.

The self-cleavage reaction catalysed by the HDV ribozyme is a transesterification reaction, which yields products with 2',3'-cyclic phosphate and 5'-hydroxyl termini. Eighty-five contiguous nucleotides are required for activity of both genomic and antigenomic sequences<sup>6</sup>. Remarkably, the presence of a single nucleotide located immediately 5' to the cleavage site is sufficient for cleavage, and the identity of this –1 nucleotide has only small effects on the reaction rate<sup>2</sup>. On the basis of sequence comparison and compensatory mutagenesis studies, Perrotta and Been<sup>7</sup> proposed closely related secondary-structure models for the genomic and antigenomic HDV ribozymes that have a pseudoknot as their distinguishing feature (Fig. 1).

The HDV ribozyme is the fastest known naturally occurring self-cleaving RNA, and also stands out among ribozymes because of its stability to denaturants and its lack of requirement for specific metal ions. The first-order rate constant for a genomic ribozyme has been estimated to be 52 reactions per min at 37 °C (ref. 8). As the optimal temperature *in vitro* is about 65 °C (ref. 9), this ribozyme can cleave itself at a rate of more than 1 per second. For comparison, the hammerhead ribozyme cleaves itself at rates of about 1 per min (refs 10, 11), and ribonuclease A cleaves UpG with a first-order rate constant of 69 per second (ref. 12). Remarkably, the HDV ribozyme is active in 5 M urea or 18 M formamide<sup>13,14</sup>. The ribozyme seems to have a nonspecific requirement for divalent cations for activity; cleavage is observed even in very low (<0.1 mM) concentrations of Ca<sup>2+</sup>, Mg<sup>2+</sup>, Mn<sup>2+</sup> or Sr<sup>2+</sup>, and at a reduced rate in the presence of several other cations<sup>15</sup>.

We have now determined the structure of the self-cleaved form of a genomic HDV ribozyme by X-ray crystallography and refined the atomic model against diffraction data extending to 2.3 Å resolution. The HDV ribozyme adopts an intricate fold that buries the active site deep within a catalytic cleft. The structure is in agreement with



**Figure 1** Pseudoknotted secondary-structure models of **a**, genomic and **b**, antigenomic HDV ribozymes proposed by Perrotta and Been<sup>7</sup>. A pseudoknot is a nucleic acid structure characterized by base-pairing between nucleotides in the loop of a conventional hairpin duplex with complementary residues outside the hairpin<sup>48</sup>. RNAs are numbered starting at the cleavage site<sup>1</sup> (large arrows). Base-paired regions are denoted P1 to P4 in a 5' to 3' direction; L3 and L4 are loops capping the corresponding paired regions. Single-stranded segments joining helices are indicated by J, followed by the helices they connect, in a 5' to 3' direction, as J1/2, J1/4 and J4/2. X denotes the nucleotide preceding the cleavage site, which can have any base. Upper-case letters denote nucleotides whose mutation results in marked reduction in ribozyme activity, and could not be rescued by compensatory mutations elsewhere. Lower-case letters mark nucleotides whose identity seems to be unimportant for ribozyme function<sup>8,9,30</sup>. Boxed nucleotides correspond to those found to be strongly (grey) or weakly (white) protected from hydroxy-radical cleavage<sup>25</sup>. Two-headed arrows mark positions of photocrosslinks formed between an azidophenacyl group placed 5' to G1 and nucleotides in J4/2 (ref. 25).



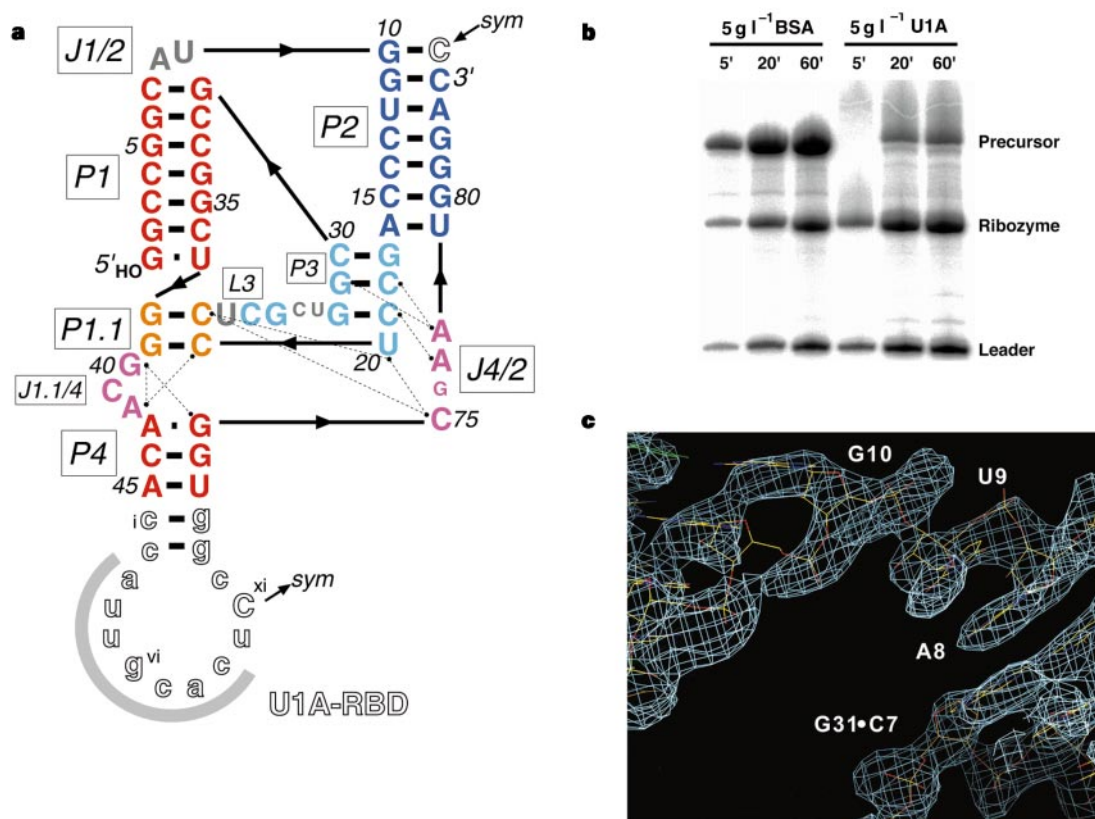
existing biochemical data, and provides a three-dimensional scaffold for understanding the mode of action of this unique RNA catalyst.

### Crystallization and structure determination

The P4 stem (Fig. 1) of the HDV ribozyme can be drastically shortened or mutagenized without detriment to the ribozyme activity, indicating that it stabilizes the active structure without participating directly in catalysis<sup>2</sup>. We reported previously that placing a crystallization module based on an RNA tertiary interaction into the P4 stem of an HDV ribozyme yielded active RNAs that crystallized readily<sup>16</sup>. To obtain well-ordered crystals, we developed a new type of crystallization module. The solvent-exposed P4 stem was engineered to contain a high-affinity binding site for the small, basic RNA-binding domain of the U1A spliceosomal protein (U1A-RBD, Fig. 2a). U1A-RBD has been extensively characterized

biochemically, and the structure of its complex with a cognate 21-nucleotide RNA stem loop has been determined previously at 1.92 Å resolution<sup>17</sup>. The alteration of P4 did not adversely affect the self-cleavage activity of the HDV ribozyme, nor did the binding of U1A-RBD (Fig. 2b). We reasoned that co-crystallization might result in better ordered crystals, because the protein–RNA complex would have a greater variety of surface functional groups than would a ‘naked’ RNA. This approach enabled crystallization of two unrelated types of ribozyme (A.R.F. and J.A.D., unpublished observations) and might be a powerful general technique for the crystallization of RNA.

As a significant proportion of HDV ribozymes transcribed *in vitro* is misfolded (ref. 2; A.R.F. and J.A.D., unpublished observations), we used self-cleaved HDV ribozymes, which have gone through the transition state, for crystallization. Biochemical<sup>2</sup> and spectroscopic<sup>14</sup> data indicate that the structures of product and precursor ribozymes



**Figure 2** Complex of the HDV ribozyme and U1A-RBD. **a**, The genomic HDV ribozyme RNA used for structure determination, rearranged to reflect aspects of the three-dimensional structure. Upper-case letters correspond to nucleotides in the wild-type HDV ribozyme, coloured to match the three-dimensional structure representations in subsequent figures. Small upper-case letters denote nucleotides that are extruded from helical stacks. Nucleotides of the U1A-binding site engineered into helix P4 (compare with Fig. 1a), and which do not participate in the ribozyme structure, are shown in outlined lower-case letters; those making close contacts with the U1A-RBD face a light grey semicircle. The numbering is as in Fig. 1a, except for the U1A-binding loop, for which the numbering is in lower-case Roman numerals. Thick short lines represent Watson–Crick base pairs; small squares indicate a wobble G–U and an A–G base pair; lines with embedded arrowheads represent covalent connections; and thin dotted lines indicate some long-range contacts found in the structure. The nucleotide Cxi forms base pairs with a symmetry-related molecule completing the 3' terminus of P2. The complementarity to J1/2 ends at C85 (Fig. 1a), defining the length of P2 of the genomic ribozyme as being seven base pairs. **b**, self-cleavage activity of the crystallized HDV ribozyme construct during transcription (see Methods) in the

presence of either 5 g l<sup>−1</sup> bovine serum albumin (BSA) or U1A-RBD (U1A), and traces of [ $\alpha$ -<sup>32</sup>P]GTP. Aliquots were removed at the indicated times, and reactions stopped by mixing with formamide and EDTA and boiling. The RNAs were resolved on an 8 M urea/12% polyacrylamide gel, and visualized by autoradiography. The build-up of product is similar in both reactions; the precursor band is smeared in the presence of U1A-RBD under these conditions. Gel mobility-shift experiments showed that both precursor and ribozyme are bound by U1A-RBD under conditions suitable for transcription (not shown). **c**, Experimental electron-density map of J1/2 contoured at one s.d. from the mean density level, superimposed on the refined atomic model. The map<sup>44</sup> was calculated using all reflections from the crystal II data set and four-wavelength MAD phases (Table 1) extended to 2.7 Å with SOLOMON<sup>43</sup>. The model shown comprises the uppermost base pairs of helices P1 and P2, and nucleotides A8 and U9 which form the J1/2 ‘bridge’. The backbone torsion angles are of A-form up to U9, and then the chain makes a sharp bend around the phosphate of G10 which, with Cxi, forms the top base pair of P2. The slight misalignment of the model and of the electron-density map reflects differences in unit-cell dimensions of the crystals used for MAD phasing and for refinement (Table 1 and Methods).

are similar, as would be expected if no nucleotides located 5' of the cleavage site participated in forming the ribozyme architecture.

Our best crystals were of the complex of a product RNA (Fig. 2a) and the selenomethionyl version of a 98-residue double-mutant U1A-RBD construct<sup>17</sup>. The structure of the complex was determined by multiwavelength anomalous diffraction (MAD) at 2.7 Å resolution, and subsequently refined at 2.3 Å to a free R factor of 28.4% (Fig. 2c and Table 1; see Methods). Experimental details, exhaustive refinement, and analysis of crystal contacts will be described elsewhere (A.R.F. and J.A.D., manuscript in preparation).

## Overview of the structure

The HDV ribozyme folds into a compact structure comprising five helical segments connected as a nested, double pseudoknot. In addition to the four biochemically identified base-paired regions, P1 to P4, the crystal structure reveals the presence of the two-base-pair helix P1.1 (Figs 2a and 3; see Figs 1 and 2a for numbering and segment name conventions). The five helical segments form two parallel stacks: P1, P1.1 and P4 stack nearly co-axially, whereas P2 and P3 form a second co-axial stack (Fig. 3). The two helical stacks are joined side-by-side by five strand-crossovers, located at the joining regions J1/2 and J4/2, the connection between P3 and P1, and at either end of P1.1. It is the unexpected formation of P1.1 by nucleotides of J1/4 and loop L3, believed until now to be single-stranded (Fig. 1), that introduces a second pseudoknot into the structure. The interactions that produce P1.1 are the vital interactions that constrain the ribozyme into its unique three-dimensional fold by bracing together the P1 substrate helix, the P2–P3 stack, the P4 helix, and the functionally important (see below) J4/2 region. No tightly bound metal ions were found in the HDV ribozyme; the structure appears to be stabilized entirely by base-pairing, stacking, and non-canonical base–backbone and backbone–backbone interactions. The compact, convoluted fold buries the 5'-hydroxyl leaving group of the self-cleavage reaction deep within an active-site cleft (Figs 3 and 4a) that is lined by nucleotides that are known to be important for catalysis.

The helices P1.1 and P3, which constitute the second pseudoknot, stack with the helices (P1 and P2) that form the pseudoknot

predicted by Perrotta and Been<sup>7</sup>. This stacking nests the P1.1–P3 pseudoknot within the P1–P2 pseudoknot. This order of stacking is different from that of classical pseudoknots, in which two consecutive stems stack co-axially and in which the two strand-crossovers interact with the major and minor grooves<sup>18</sup>. The three-dimensional fold of the nested double pseudoknot probably occurs in other RNAs. For example, such a structure was proposed for the translational repressor of the  $\alpha$ -operon of *Escherichia coli*<sup>19</sup>. The thermal unfolding of the two inner helices of the  $\alpha$ -repressor (these inner helices correspond to P2 and P3 in the HDV ribozyme) is coupled<sup>20</sup>, indicating that they might be stacked like P2 and P3 are in the HDV ribozyme. Another double-pseudoknot fold has been proposed for a very active RNA ligase identified by *in vitro* selection<sup>21</sup>.

A computer graphics model of the HDV ribozyme<sup>8</sup> did not predict the presence of P1.1 or, therefore, the unique, doubly pseudoknotted fold of the ribozyme. Because that model lacked the structural constraints imposed by P1.1, P4 was rotated roughly 180° away along its helical axis from the correct confirmation. As a result, the active site of this model bears no resemblance to the active site seen here in the crystal structure. Two NMR structure determinations of RNA sequences corresponding to isolated P2–P3–L3 segments have been reported<sup>22,23</sup>. As the base-pairing interactions of L3 with J1/4 are absent in these constructs, the two NMR structures exhibit different base-pairing schemes from those seen in the crystal structure.

If an RNA fold is defined as a three-dimensional arrangement of at least two separate helical stacks, the HDV ribozyme represents the fourth RNA fold whose structure is known; the other three are transfer RNA, the hammerhead ribozyme, and the P4–P6 domain of the *Tetrahymena* group I intron. In tRNA, the two helical stacks are perpendicular, whereas in the hammerhead ribozyme they join in one region to form a Y-shaped molecule. The HDV ribozyme superficially resembles the P4–P6 domain in that two helical stacks are arranged side by side (Fig. 3c). However, the 160-nucleotide P4–P6 domain is not pseudoknotted, and its two helical stacks are brought together by two sets of loop-to-helix-groove contacts involving non-canonical interactions (the A-rich-bulge–P4 and tetraloop–receptor contacts). The smaller HDV ribozyme comprises

**Table 1 Crystallographic statistics**

| Diffraction data                                           |                        |                |                                           |                                         |                                                                       |                                                      |
|------------------------------------------------------------|------------------------|----------------|-------------------------------------------|-----------------------------------------|-----------------------------------------------------------------------|------------------------------------------------------|
| Data set                                                   | Wavelength (Å)         | Resolution (Å) | Number of reflections*<br>observed/unique | Data coverage (%)<br>overall/last shell | $\langle I \rangle / \langle \sigma(I) \rangle$<br>overall/last shell | $R_{\text{sym}}^{\dagger}$ (%)<br>overall/last shell |
| Crystal I                                                  | 0.9761 ( $\lambda_1$ ) | 41.3–2.9       | 55,889/17,817                             | 95.7/97.6                               | 21.7/5.2                                                              | 6.5/31.8                                             |
|                                                            | 0.9794 ( $\lambda_2$ ) | 42.5–2.9       | 85,008/18,459                             | 99.1/99.9                               | 26.2/6.5                                                              | 7.8/39.3                                             |
|                                                            | 0.9792 ( $\lambda_3$ ) | 35.3–2.9       | 56,668/17,962                             | 96.4/98.0                               | 21.0/3.6                                                              | 6.8/41.1                                             |
| Crystal II                                                 | 1.1390 ( $\lambda_4$ ) | 41.5–2.7       | 68,210/21,826                             | 93.5/93.9                               | 17.7/4.4                                                              | 9.0/34.7                                             |
| Crystal III                                                | 0.9209                 | 28.6–2.3       | 72,648/18,914                             | 95.5/96.5                               | 25.9/4.1                                                              | 6.6/41.4                                             |
|                                                            |                        |                |                                           |                                         |                                                                       |                                                      |
| MAD analysis (crystals I and II)                           |                        |                |                                           |                                         |                                                                       |                                                      |
| Phasing power $^{\ddagger}$ (acentric reflections)         |                        |                |                                           |                                         |                                                                       |                                                      |
| Resolution range (Å)                                       | $\lambda_1$ ano        |                | $\lambda_2$ iso/ano                       |                                         | $\lambda_3$ iso/ano                                                   | $\lambda_4$ iso/ano                                  |
| 42.5–2.9                                                   | 2.83                   |                | 2.39/2.64                                 |                                         | 2.22/3.10                                                             | 0.76/0.58                                            |
| 3.2–2.9                                                    | 1.41                   |                | 1.65/1.39                                 |                                         | 1.37/1.44                                                             | 1.04/0.21                                            |
| $R_{\text{krout}}^{\S}$ (acentric reflections, %)          |                        |                |                                           |                                         |                                                                       |                                                      |
| Resolution range (Å)                                       | $\lambda_1$ iso/ano    |                | $\lambda_2$ iso/ano                       |                                         | $\lambda_3$ iso/ano                                                   | $\lambda_4$ iso/ano                                  |
| 42.5–2.9                                                   | 1.7/4.3                |                | 2.5/3.8                                   |                                         | 2.9/4.6                                                               | 9.1/4.0                                              |
| 3.2–2.9                                                    | 6.1/13.1               |                | 6.2/12.0                                  |                                         | 8.7/16.0                                                              | 5.9/11.7                                             |
| Mean figure of merit $^{\parallel}$ (acentric reflections) |                        |                | Resolution range (Å)                      |                                         |                                                                       |                                                      |
|                                                            |                        |                | 42.5–2.9                                  |                                         |                                                                       |                                                      |
|                                                            |                        |                | 3.1–2.9                                   |                                         |                                                                       |                                                      |
|                                                            |                        |                | 0.635                                     |                                         |                                                                       |                                                      |
|                                                            |                        |                | 0.425                                     |                                         |                                                                       |                                                      |

\* Anomalous reflections counted separately for crystals I and II.

$^{\dagger} R_{\text{sym}} = \sum_i |I_i - \langle I \rangle| / \sum_i I_i$ , where  $I_i$  is the observed intensity and  $\langle I \rangle$  is the statistically weighted average intensity of multiple measurements of symmetry-related reflections.

$^{\ddagger}$  Phasing power =  $\langle |F_{\text{H}}| \rangle / \langle |F_{\text{H}}| - |F_{\text{P}} + F_{\text{H}}| \rangle$ , where  $\langle |F_{\text{H}}| - |F_{\text{P}} + F_{\text{H}}| \rangle$  is the residual lack-of-closure error; 'iso' and 'ano' refer to contributions from isomorphous and anomalous differences.

$^{\S} R_{\text{krout}} = \sum_i |F_{\text{H}}| - |F_{\text{P}} + F_{\text{H}}| / \sum_i |F_{\text{H}}|$ .

$||$  Mean figure of merit =  $\langle \sum_i P(\alpha) e^{i\alpha} / \sum_i P(\alpha) \rangle$ , where  $\alpha$  is the phase and  $P(\alpha)$  is the phase-probability distribution.

mainly standard Watson–Crick base pairs (Fig. 2a), and achieves its compact structure by adopting its intricate double-pseudoknot fold.

The convoluted fold of the HDV ribozyme appears to contribute significantly to its stability, as disruption of either of its two pseudoknots results in a marked loss of activity. Cleavage of the ribozyme at J1/2, which deletes the first pseudoknot, produces molecules that are active *in trans*. However, those molecules have cleavage rates that are about two orders of magnitude slower than the rates of their parent *cis*-ribozymes, and, unlike the intact ribozymes, are inactivated by denaturants<sup>2</sup>. Mutagenesis has been used to demonstrate the importance of the four conserved nucleotides of P1.1 for ribozyme function. Their deletion, or their mutation into nucleotides that would disrupt the P1.1 helix and hence delete the second pseudoknot, reduced ribozyme activity by three to five orders of magnitude<sup>8,24</sup>.

### Structural components of the active site

Hydroxy-radical footprinting of genomic and antigenomic ribozymes (Fig. 1) showed strong protection of the RNA backbone on the 5' side of P3 and L3, of the nucleotides at the junction of P3 and P1, of J4/2, and weaker protection of guanines 38–40 (genomic numbering)<sup>25</sup>. In good agreement with this footprinting pattern, the crystal structure shows that these segments line a buried cleft around the 5'-hydroxyl leaving group, resulting from the self-scission reaction (Figs 3 and 4a). The position of this functional group identifies the active site of the HDV ribozyme. Three sides of this active-site cleft are formed by the substrate-bearing P1 helix, J4/2, and a niche formed by the minor groove of P3 and L3. The cleft is enclosed on top by the arching RNA backbone in the junction of P3 and P1, and supported from below by the P1.1 helix.

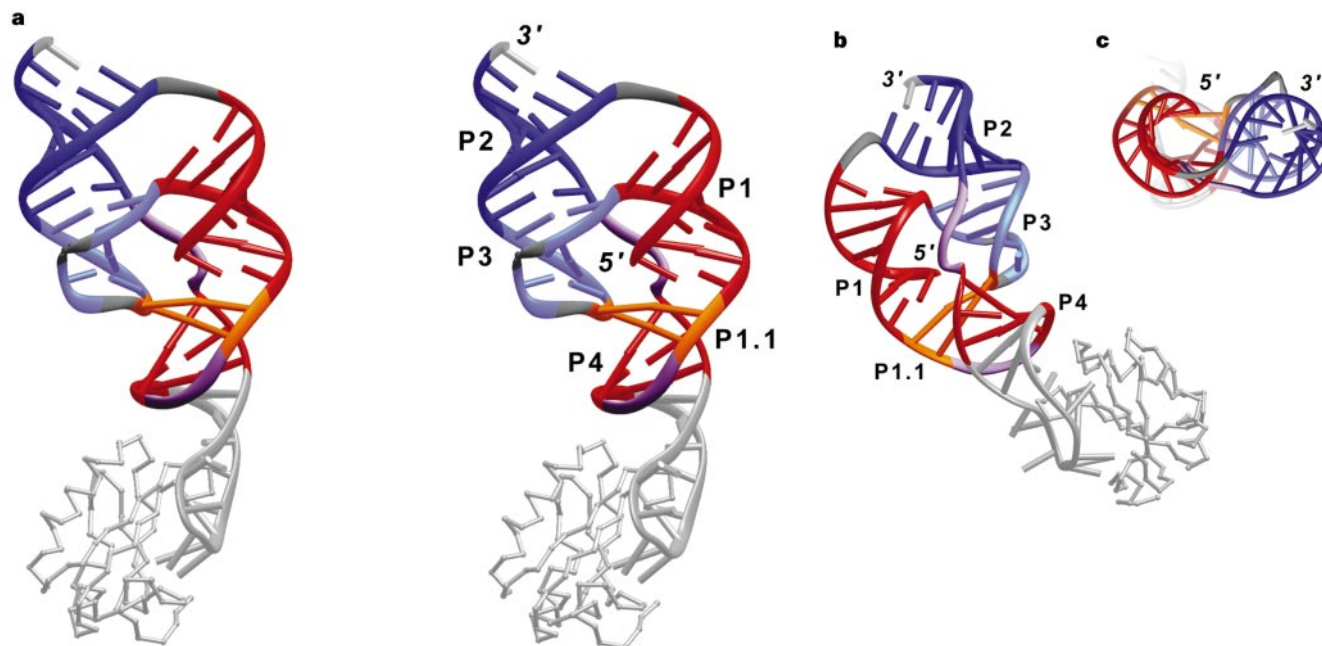
**Positioning of the substrate-bearing helix.** The substrate-bearing P1 helix is positioned in the active site of the HDV ribozyme both by a pair of helical crossovers at its top and by the stacking of a G-U wobble (G1·U37) on P1.1 at its bottom (Figs 2a, 3 and 4b).

Functional groups in the grooves or the phosphate backbone of the helix do not make lateral contacts with the rest of the ribozyme. Site-directed mutagenesis showed that, provided that base-pairing and total helix length were maintained, the composition of P1 could be varied freely, except for the G1·U37 wobble pair<sup>2</sup>. The structure shows that because of the co-axial stacking of P1 on P1.1, changing the length of P1 would place the crossovers out of register and destabilize the ribozyme. The mutagenesis studies indicated that both stacking energy and placement of the 5'-hydroxyl group (which bears the scissile phosphate in the precursor) resulting from the wobble pair are probably essential for activity, as A·C or G·C, but not any other standard or non-canonical pairs, could partially replace the G·U wobble<sup>26</sup>.

The structure shows that, unlike in the group I intron, in which the cleavage site is recognized in part through the N2 amino group of guanosine exposed as a result of wobble base pairing<sup>27</sup>, the HDV ribozyme exploits the backbone distortion inherent in the G·U wobble. The 5'-hydroxyl group of G1 is pushed deeper towards the centre of the ribozyme because of the twist of G1 into the minor groove and of U37 into the major groove (Figs 3c and 4b). In addition, the helix twist, together with a displacement of the helix axes of P1 and P1.1, results in G1 stacking onto G38 at the top of the opposite strand of P1.1 (Figs 3a and 4b).

**A strand-crossover overhangs the core.** The P3–P1 strand-crossover arches over the active site (Figs 3a and 4c). The 3' side of P3 exits the P2–P3 stack at C30. Without any spacer nucleotides, the backbone makes a sharp change in direction about the phosphate of G31, which then participates in the uppermost base pair of P1. The P2–P3 stack is continued by the incursion of the strand coming from J4/2, which forms the 3' side of P3. The strong hydroxy-radical protection of C30 and G31 (ref. 25) (Fig. 1) might result from the riboses of these two nucleotides facing an enclosed cavity formed by P1, P2 and the J1/2 bridge (Figs 2c and 3).

A triple-base interaction stabilizes this P3–P1 strand-crossover (Fig. 4c): A78, the last nucleotide of J4/2, forms four hydrogen

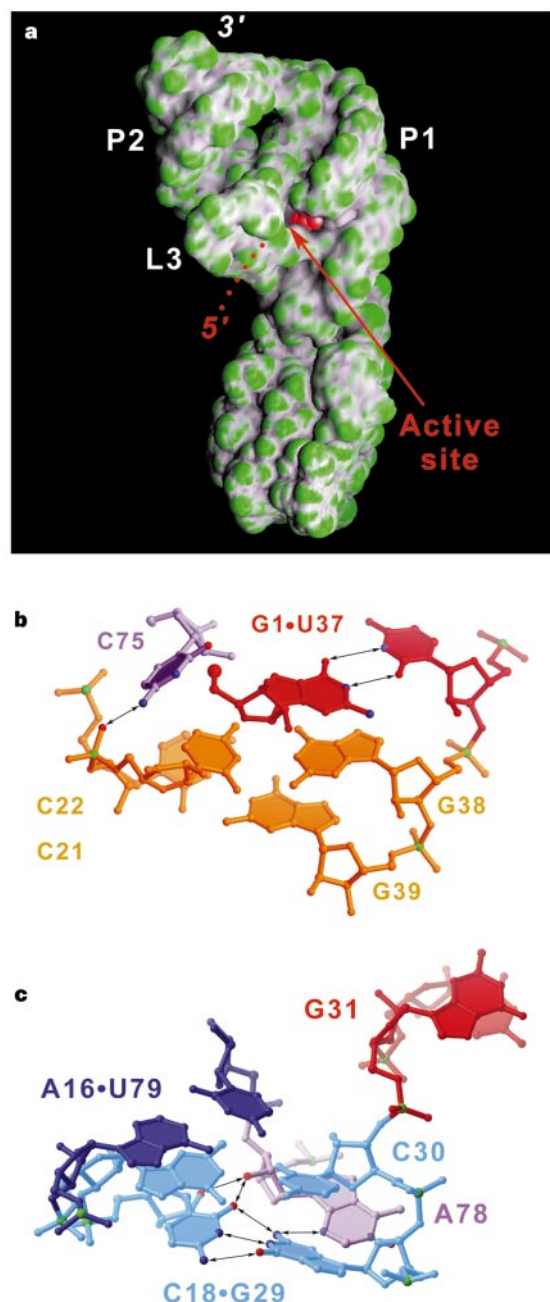


**Figure 3** Ribbon and stick representations of the crystal structure of the HDV ribozyme, colour-coded as in Fig. 2a. The path of the RNA backbone is shown as a ribbon, and base-paired nucleotides as sticks. The unpaired nucleotides from J1/2 (grey), J1.1/4 and J4/2 (pink) and L3 (grey) are omitted for clarity. The U1A-RBD protein is represented as an  $\alpha$ -carbon ball-and-stick model. Paired regions and

RNA termini are labelled. Cxi at the 3' terminus is in white. **a**, Stereoview of the front of the structure. **b**, Back view of the molecule, tilted to emphasize the snug fit of the J4/2 region between the two helical stacks, and the intricate pseudoknot crossovers. **c**, Top view of the structure, showing that the two helical stacks are nearly parallel to each other.



bonds with the minor groove face of the penultimate pair of P3 (C18-G29). This base triplet has been observed before, both at the P5abc junction of the P4–P6 domain<sup>28</sup>, and as part of a base-quadruplet interaction between GAAA tetraloops and either an intramolecular tetraloop receptor<sup>28</sup> or an RNA stem in a crystal



**Figure 4** Structural components of the active site. **a**, Solvent-accessible surface of the ribozyme–U1A–RBD complex, orientated as in Fig. 3a, colour-coded by its curvature (green, convex; grey, concave<sup>49</sup>). The atoms of the ribose ring of G1 are shown as red spheres; the 5′-hydroxyl group points away from the reader into the ribozyme. The trajectory proposed to be followed by nucleotides located 5′ of the cleavage site in the precursor is indicated by the row of small red dots. **b**, The substrate-bearing nucleotide G1 forms a wobble pair with U37 which stacks on P1.1. The 5′-hydroxyl group (shown as a larger sphere) is within hydrogen-bonding distance of the Watson–Crick face of C75. Nucleotides are coloured as in Figs 2 and 3. Phosphorus atoms are coloured green. Oxygen and nitrogen atoms that participate in noteworthy hydrogen bonding (denoted by two-headed arrows) are coloured red and blue, respectively. **c**, Triple-strand junction between P3 and P1, orientated as in Fig. 3a. A hydrogen bond between the N1 group of A78 and the O2′ group of G29 is obscured in this view.

contact<sup>29</sup>. This interaction explains, in part, the strict requirement for A78 (refs 8, 30). The exocyclic N6 of A78 is not involved in hydrogen bonding and points towards the active site, where it would be available for coordination, and may participate in catalysis.

**P3 and L3 form a niche.** P3 is composed of three base pairs as predicted<sup>7</sup> (Figs 3 and 4c). Consistent with mutagenesis studies<sup>2,9,31</sup>, P3 makes no sequence-specific contacts with the rest of the ribozyme. The minor groove face of P3 and the unpaired nucleotides U20, C24 and G25 form a niche that flanks the active site (Figs 3, 4a and 5a). The helical stack of P3, but not the base pairing, is continued by U20 and G25 below the C19–G28 pair. U20 and G25 are too far from each other to form any hydrogen bonds between their Watson–Crick faces. Mutagenesis of these two nucleotides to sequences that would force pairing (such as G25A) resulted in ribozymes that were 3,000 times less active<sup>8</sup>. The P2–P3 helical stack ends with C24 stacking underneath G25, where it makes a hydrogen bond between its N4 and the pro-*S*<sub>p</sub> phosphate of C22 (Fig. 5a). Of the remaining nucleotides in L3, C21 and C22 are part of P1.1, U23 lies underneath C24, and C26 and U27 extrude from the helical stack, in good agreement with chemical-modification experiments<sup>32</sup>. The pyrimidine bases of C26, U27 and, to a lesser extent, U23 (grey in Fig. 3) are not well-ordered in our crystals (see Methods). Results of biochemical experiments indicate that these three mobile nucleotides act mainly to connect catalytically important segments of P3 and L3 and that their bases have only modest structural and energetic roles<sup>9,32–34</sup>.

**P1.1 and a pedestal support the core.** Two strand-crossovers corresponding to the second pseudoknot occur at either end of P1.1, and form part of a pedestal that rests upon P4 and supports both the P2–P3–L3 stack and the cleavage-site G–U wobble (Fig. 5b). The non-canonical A43–G74 base pair at the top of P4 (Fig. 5d) occurs at the beginning of the P4 helix, and results in an opening of the minor groove. The pedestal is completed by novel interactions around J1.1/4. The C21–G39 pair at the bottom of P1.1 stacks on the 3′ side of the A43–G74 pair only, because of the large (100°) twist angle between these pairs. The J1.1/4 segment spanning G39–A43 (Figs 3 and 5b) adopts a unique turn conformation that is different from the well-characterized U- or  $\pi$ -turns. G40, the first of the turn nucleotides, stacks under G39, and also forms hydrogen bonds with A42 and G74. The backbone then buckles, pushing the unpaired C41 into a narrowed major groove and unstacking its bottom face. The backbone turns again, now pushing the base of A42 into the minor groove, where it stacks on A43 and buttresses C21 by hydrogen bonding (Fig. 5b). Beyond this, the chain enters P4 where it adopts an A-conformation. The turn breaks the helical stack between P1.1 and P4 on one strand but not on the other, connecting P1 and P4 while also reinforcing the minor groove of P1.1.

**A turn and ribose zipper buttress the core.** The single-stranded region J4/2 closes the rear side of the active site by connecting P4 with P2, and incorporates a ‘trefoil’ turn (Fig. 5c, d) that positions important nucleotides<sup>2,8</sup> in the active site, arranging G74, C75 and G76 as the three leaflets of a clover and fully extruding the base of G76 into solvent. C75 projects deep into the core of the ribozyme, and is held in place by a hydrogen bond between its N4 and the pro-*R*<sub>p</sub> oxygen of C22, as well as by stacking interactions with A77 and backbone interactions. The top two nucleotides of J4/2, namely A77 and A78, are held in place by the base-triplet interaction in the P3–P1 junction (Fig. 4c), and are also involved in a two-tiered ribose-zipper<sup>28</sup> interaction with C19 and C18, respectively (Fig. 5d). Consistent with their marked protection from hydroxy-radical attack<sup>25</sup> (Fig. 1), chemical modification<sup>32</sup> and enzymatic digestion<sup>35</sup> the riboses of C75–A78 are buried by the conformation of J4/2 and the dense network of backbone hydrogen bonds.

As expected from its position within the active site next to the 5′-hydroxyl leaving group produced by the self-scission reaction (Figs 4b and 5a), C75 cannot be mutated to any other nucleotide

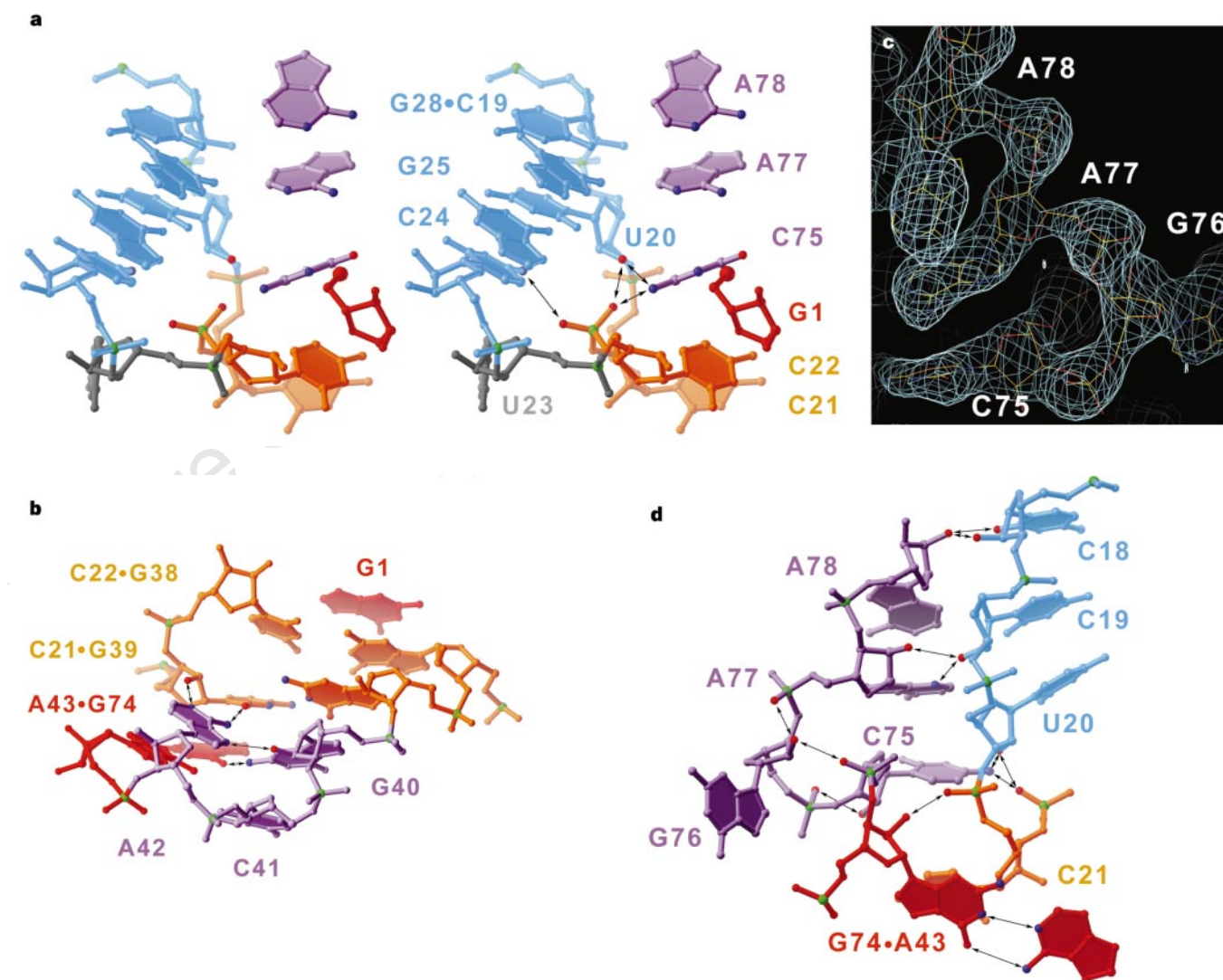
without completely abolishing ribozyme activity<sup>8</sup>. However, the extruded G76 can be mutated to uridine with only modest loss of activity<sup>8</sup>; furthermore, results of a modification-interference study showed that opening the purine ring of this guanosine resulted in an increase in ribozyme activity<sup>33</sup>. Both A77 and A78 are also sensitive to changes<sup>8</sup>.

### Active site and catalysis

Together with existing biochemical information, the crystal structure of the HDV ribozyme offers some hints to the chemical mechanism of this RNA catalyst. A solvent-accessible surface representation of the crystal structure (Fig. 4a) shows that the 5'-hydroxyl group, which marks the location of the active site, lies buried at the bottom of a deep cleft, surrounded by the structural elements described above. Mutations in all of these elements are deleterious to the function of the ribozyme, indicating that the product of the self-scission reaction has a structure similar to that of the precursor and the transition state. Consistent with a transesterification mechanism in which the 2'-hydroxyl group of the -1 nucleotide performs a nucleophilic attack on the phosphorus group,

substitution of the -1 ribose with a 2'-deoxyribose abolished the reaction<sup>2</sup>. Thus, structural elements present in the product RNA whose structure we have determined must position the -1 ribose for nucleophilic attack, provide a base with which to activate its 2'-hydroxyl group, and perhaps stabilize the transition state by neutralizing the developing negative charge in the leaving group.

Our structure of the HDV ribozyme and the available biochemical data do not reveal how metal ions might participate in catalysis. The rate of cleavage of the ribozyme does not show a marked pH dependence<sup>5,36</sup>. Phosphorothioate substitution at the cleavage site showed that the pro-R<sub>p</sub> oxygen cannot be substituted by a sulphur without loss of activity<sup>37</sup>; cleavage was not recovered by the addition of Mn<sup>2+</sup>. The only other phosphate oxygen sensitive to sulphur substitution was the pro-R<sub>p</sub> oxygen of C22. This substitution was also insensitive to Mn<sup>2+</sup>. This pro-R<sub>p</sub> oxygen forms a hydrogen bond with the *trans* face of the N4 amino group of C75 (Figs 4b and 5a, d), so the phosphorothioate substitution probably results in steric clash. The lack of requirement for a specific metal ion indicates that the ion does not participate directly in activating the nucleophile, but instead acts as a charge sink.



**Figure 5** Structural features of the active site. **a**, The P3-L3 niche cradles the active site. The stereoview shows how the phosphate ribose backbone crosses over from the P3-L3 stack to P1.1 and back, passing through the unstacked nucleotide U23. The ribose of G1 and the bases of the stacked nucleotides from J4/2 are shown. **b**, The P1.1 helix and the pedestal. **c**, A  $2F_o - |F_c|$  map<sup>44,45</sup>, with data from 20 to 2.3 Å, contoured at 1 s.d., showing the nucleotides of

J4/2 as seen from the major groove of P1. **d**, The trefoil turn (orientated roughly as in Fig. 3b) is the sharp twist in the RNA chain between G74 and A77. This is stabilized by a complex network of interactions between O2' groups and phosphate oxygens, a two-tiered ribose zipper, and base-backbone hydrogen bonds. Figures 3, 4b, c and 5a, b, d were prepared with RIBBONS<sup>50</sup>.

In a *trans* ribozyme, an azidophenacyl group placed at the cleavage site formed crosslinks with residues at the 3' side of L3 (equivalent to G25 and C26)<sup>25</sup>. Photocrosslinking of a cleavable *trans* ribozyme substituted with deoxythiouridine at position -2 resulted in crosslinks from this deoxythiouridine to the nucleotide corresponding to C75. Deoxythiouridines at positions -2 and -1 in non-cleavable ribozymes formed crosslinks with nucleotides corresponding to C21 and G25 in the L3 niche<sup>38</sup>. We propose that the RNA chain leading to the scissile bond enters the core of the HDV ribozyme through the niche formed by these mutation-sensitive nucleotides (Figs 4a and 5a). If so, the phosphodiester backbone must make a sharp bend at or near the scissile bond, so that it can adopt the helical conformation it has in P1. Qualitatively, it appears that the parts of bases in the niche that form minor grooves would be well placed to position the ribose for attack. The extrusion of U23 from the base stack creates a pocket at the bottom of L3 in which the base of nucleotide -1 could be accommodated without making specific interactions (Fig. 5a).

From results of crosslinking and mutagenesis studies, and because of its proximity to the 5'-hydroxyl group in our structure, it seems that C75 is probably important in catalysis. As metal ions do not appear to participate directly in catalysis, we propose that C75 acts as the general base that activates the 2'-hydroxyl group of nucleotide -1 for nucleophilic attack. The *trans* face of the N4 amino group of C75 participates in an interesting network of hydrogen-bond interactions (Fig. 5a, d). It is positioned so that it could donate a hydrogen bond to either the O2' group of U20, or the pro-R<sub>p</sub> phosphate oxygen of C22. The O2' of U20 probably donates a hydrogen bond to the same phosphate oxygen. There are, therefore, three polar ligands arranged at the vertices of a triangle with sides of length ~2.8 Å and only two hydrogens (Fig. 5a, d). This, and the negative electrostatic potential in the region of C75 (resulting from the trefoil turn), could perturb the pK<sub>a</sub> values of C75, making it basic.

In summary, the crystal structure of a genomic HDV ribozyme reveals a unique, nested double-pseudoknot fold that allows this catalytic RNA to form a deep, solvent-inaccessible active-site cleft, much like that of protein enzymes. The only other ribozyme for which a structure is available, the hammerhead ribozyme, places its scissile bond in a solvent-exposed portion of a typical A-form helix<sup>10,11</sup>. Our structure of the HDV ribozyme explains the available biochemical data, and will act as the starting point for detailed studies of the chemical mode of action of this very active catalytic RNA. □

## Methods

**RNA and protein preparation.** Plasmid pDU9 is a derivative of pUC19 that encodes the HDV ribozyme we studied, preceded by a 41-nucleotide leader sequence and a T7 RNA polymerase promoter. The insert was prepared by the polymerase chain reaction from overlapping synthetic DNA oligonucleotides, and is followed by a *BsaI* restriction site, such that run-off transcription of the linear plasmid followed by ribozyme self-cleavage produces the 72-nucleotide HDV ribozyme RNA (Fig. 2a). The RNA has homogeneous 5', and heterogeneous 3' termini, because of addition of a few random nucleotides by the polymerase. Attempts at producing a homogeneous 3' end failed. Transcription and purification were done as described<sup>16</sup>.

Selenomethionyl U1A-RBD was produced in the methionine auxotroph *Escherichia coli* strain B834 (Novagen), transformed with the (A1-98 Y31H, Q36R) T7 expression plasmid<sup>17</sup> (provided by K. Nagai) and grown in minimal medium supplemented with selenomethionine. Cell lysate was clarified by centrifugation, and the supernatant fractionated with polyethyleneimine and ammonium sulphate. The protein was further purified by cation-exchange (SP-Sephacrose FF, Pharmacia), size-exclusion (Superdex-75 PG, Pharmacia), and hydroxyapatite (CHT-I, BioRad) chromatography.

**Crystallization and data collection.** Annealed RNA was mixed with one molar equivalent of selenomethionyl U1A-RBD protein to yield a complex concentration of 0.5 mM (in 1.25 mM MgCl<sub>2</sub>). For crystallization by the

sitting-drop vapour-diffusion technique, 4 µl complex solution was mixed with 2 µl of a reservoir solution consisting of 12.5–14% (w/v) polyethyleneglycol monomethyl ether of average molecular weight 2,000 (PEG-MME 2000), 100 mM Tris-HCl, pH 7.0, 200–250 mM Li<sub>2</sub>SO<sub>4</sub> and 4 mM spermine-HCl. In addition, reservoir solutions contained 0.2 mM cobalt(III) hexamine chloride (CoHex, crystals I and III); 0.4 mM osmium(III) hexamine chloride triflate (OsHex, crystal II); and 5 mM MgCl<sub>2</sub> (crystal III). Trigonal trapezohedra grew at 25 °C over several weeks, and attained maximum dimensions of 0.8 × 0.6 × 0.6 mm<sup>3</sup> (space group *R*32,  $\alpha = 90^\circ$ ,  $\gamma = 120^\circ$ ). Cell dimensions were  $a = 108.7$  Å,  $109.15$  Å and  $109.35$  Å, and  $c = 190.38$  Å,  $190.93$  Å and  $190.68$  Å, for crystals I, II and III, respectively. For cryogenic data collection, crystals were transferred to solutions containing 25% (w/v) PEG-MME 2000, 100 mM TRIS-HCl pH 7.0, 250 mM Li<sub>2</sub>SO<sub>4</sub>, 6 mM (crystals I and II) or 25 mM (crystal III) spermine-HCl, 0.4 mM or 2.0 mM CoHex (crystals I and III, respectively), or 0.5 mM OsHex (crystal II), 2.5 mM, 2.0 mM or 10 mM (crystals I, II and III, respectively) MgCl<sub>2</sub>, 5% (v/v) PEG 6000, and 3–4% (v/v) (2*R*,3*R*)-(–)-2,3-butanediol. Crystals were mounted and flash-cooled as described<sup>16</sup>.

MAD data from crystals I and II were collected at beamline X4A of the National Synchrotron Light Source using an Raxis-IV imaging plate area detector (Rigaku) and the inverse-beam mode. Data from crystal III was obtained at beamline F-1 of the Cornell High Energy Synchrotron Source on a Quantum-4 CCD detector (Area Detectors Systems) (Table 1). Oscillation photographs were processed using the programs DENZO and SCALEPACK<sup>39</sup>, and the CCP4 suite<sup>40</sup>.

**Phase determination and structure refinement.** Four selenium sites were located in crystal I using the package SOLVE<sup>41</sup> (www.solve.lanl.gov). Density modification<sup>40</sup> of three-wavelength SOLVE phases produced a 2.9 Å electron-density map with clear protein and RNA features. Fourier maps calculated using these phases and anomalous differences from data from crystal II did not contain significant peaks except for the four selenium sites, indicating that the OsHex was not bound specifically. Selenium parameters were further refined and phase distributions and centroids were calculated with SHARP<sup>42</sup> using data collected at three wavelengths from crystal I (high-energy remote, rising inflection point, and peak, respectively), and one from crystal II (low-energy remote, Table 1). Three of the four seleniums are well-ordered; the site corresponding to Met 97 is not. Density modification and phase extension to 2.7 Å (SOLOMON<sup>43</sup>) produced an electron-density map into which most of the protein and nucleic acid residues could be built unambiguously, using program O (ref. 44). Correct registration of both macromolecule chains was confirmed by superimposing the methionine sites of the U1A-RBD–RNA 21-mer complex coordinates<sup>17</sup> on the selenium sites. The U1A–RNA 21-mer structure<sup>17</sup> was not used as a source of phase information at any stage.

Rounds of manual rebuilding, interspersed with positional, torsion-angle-simulated annealing, and tightly restrained individual *B*-factor refinement with CNS<sup>45</sup>, produced a model with an *R*-free value of 38.9% to 2.7 Å. This model was further refined against all crystal III amplitude data between 20 and 2.3 Å (17,042 and 1,861 reflections in the work and free sets, respectively), and against experimental phase-probability distributions from SHARP, using the phased maximum-likelihood target, purely repulsive van der Waals parameters, and a bulk solvent mask<sup>45</sup>. Use of the phased maximum-likelihood target<sup>45</sup> resulted in a *R*-free factor typical of a well refined model at this resolution<sup>46</sup>, with negligible overfitting ( $R_{\text{free}} = 28.4\%$ ,  $R_{\text{work}} = 28.0\%$ ). Despite the high average *B*-factors (44.2 and 91.7 Å<sup>2</sup>, respectively, for protein and RNA atoms), electron density for the RNA is of good quality (Figs 2c and 5c). In a  $\sigma A$ -weighted  $2|F_o| - |F_c|$  map<sup>45</sup> contoured at 0.9 s.d., there is a total of seven minor breaks in the RNA backbone, in P1, P2 and P4, and two larger breaks in the backbone of L3, in which electron density for the extruded bases of residues U23, C26 and U27 is also weak. Electron density for the protein moiety is good throughout, with continuous backbone density from residues 7 to 96 in the same map contoured at 2.5 s.d. All nucleotides are in *anti* conformation. Ribose puckers are all C3'-endo except for nucleotides 22, 41, 74, 75, 76, 79, v, vii, and x, which were restrained to C2'-endo. The present model consists of all RNA residues, U1A–RBD residues 4–98 (H31, R36 and S71 are all in two conformations), two sulphate and three magnesium ions, and nineteen water molecules (2,366 non-hydrogen atoms), and has root-mean-square (r.m.s.) deviations from the ideal of 0.009 Å and 1.43° for bond lengths and angles,



respectively. The r.m.s. deviation of the B-factor of covalently bonded atoms is  $0.47 \text{ \AA}^2$ . PROCHECK<sup>47</sup> analysis shows no outliers in the Ramachandran plot, and an overall G-factor better (2.7 bandwidths from the mean) than expected at this resolution.

Received 17 July; accepted 7 September 1998.

- Lai, M. M. The molecular biology of hepatitis delta virus. *Annu. Rev. Biochem.* **64**, 259–286 (1995).
- Been, M. D. & Wickham, G. S. Self-cleaving ribozymes of hepatitis delta virus RNA. *Eur. J. Biochem.* **247**, 741–753 (1997).
- Kuo, M. Y.-P., Sharmeen, L., Dinter-Gottlieb, G. & Taylor, J. Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *J. Virol.* **62**, 4439–444 (1988).
- Sharmeen, L., Kuo, M. Y.-P., Dinter-Gottlieb, G. & Taylor, J. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J. Virol.* **62**, 2674–2679 (1988).
- Wu, H.-N. *et al.* Human hepatitis delta virus RNA subfragments contain an autocleavage activity. *Proc. Natl Acad. Sci. USA* **86**, 1831–1835 (1989).
- Perrotta, A. T. & Been, M. D. The self-cleaving domain from the genomic RNA of hepatitis delta virus: sequence requirements and the effects of denaturant. *Nucleic Acids Res.* **18**, 6821–6827 (1990).
- Perrotta, A. T. & Been, M. D. A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. *Nature* **350**, 434–436 (1991).
- Tanner, N. K. *et al.* A three-dimensional model of hepatitis delta virus ribozyme based on biochemical and mutational analyses. *Curr. Biol.* **4**, 488–498 (1994).
- Thill, G., Vasseur, M. & Tanner, N. K. Structural and sequence elements required for the self-cleaving activity of the hepatitis delta virus ribozyme. *Biochemistry* **32**, 4254–4262 (1993).
- Pley, H. W., Flaherty, K. M. & McKay, D. B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68–74 (1994).
- Scott, W. G., Finch, J. T. & Klug, A. The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 991–1002 (1995).
- Richards, F. M. & Wyckoff, H. W. In *The Enzymes* (ed. Boyer, P. D.) 647–806 (Academic, New York, 1971).
- Rosenstein, S. P. & Been, M. D. Self-cleavage of hepatitis delta virus genomic strand RNA is enhanced under partially denaturing conditions. *Biochemistry* **29**, 8011–8016 (1990).
- Duhamel, J. *et al.* Secondary structure content of the HDV ribozyme in 95% formamide. *Nucleic Acids Res.* **24**, 3911–3917 (1996).
- Suh, Y.-A., Kumar, P. K. R., Taira, K. & Nishikawa, S. Self-cleavage activity of the genomic HDV ribozyme in the presence of various divalent metal ions. *Nucleic Acids Res.* **21**, 3277–3280 (1993).
- Ferré-D'Amaré, A. R., Zhou, K. & Doudna, J. A. A general module for RNA crystallization. *J. Mol. Biol.* **279**, 621–631 (1998).
- Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. & Nagai, K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438 (1994).
- Kolk, M. H. *et al.* NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA. *Science* **280**, 434–438 (1998).
- Tang, C. K. & Draper, D. E. Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. *Cell* **57**, 531–536 (1989).
- Gluck, T. C. & Draper, D. E. Thermodynamics of folding a pseudoknotted mRNA fragment. *J. Mol. Biol.* **241**, 246–262 (1994).
- Eklund, E. H., Szostak, J. W. & Bartel, D. P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**, 364–370 (1995).
- Kolk, M. H., Heus, H. A. & Hilbers, C. W. The structure of the isolated, central hairpin of the HDV antigenomic ribozyme: novel structural features and similarity of the loop in the ribozyme and free in solution. *EMBO J.* **16**, 3685–3692 (1997).
- Lynch, S. R. & Tinoco, I. Jr The structure of the L3 loop from the hepatitis delta virus ribozyme: a *syn* cytidine. *Nucleic Acids Res.* **26**, 980–987 (1998).
- Perrotta, A. T. & Been, M. D. Core sequences and a cleavage site wobble pair required for HDV antigenomic ribozyme self-cleavage. *Nucleic Acids Res.* **24**, 1314–1321 (1996).
- Rosenstein, S. P. & Been, M. D. Hepatitis delta virus ribozymes fold to generate a solvent-inaccessible core with essential nucleotides near the cleavage site phosphate. *Biochemistry* **35**, 11403–11413 (1996).
- Nishikawa, F., Fauzi, H. & Nishikawa, S. Detailed analysis of base preferences at the cleavage site of a trans-acting HDV ribozyme: a mutation that changes cleavage site specificity. *Nucleic Acids Res.* **25**, 1605–1610 (1997).
- Strobel, S. A. & Cech, T. R. Minor groove recognition of the conserved G-U pair at the *Tetrahymena* ribozyme reaction site. *Science* **267**, 675–679 (1995).
- Cate, J. H. *et al.* Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678–1685 (1996).
- Pley, H. W., Flaherty, K. M. & McKay, D. B. Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372**, 111–113 (1994).
- Kumar, P. K. R. *et al.* Random mutations to evaluate the role of bases at two important single-stranded regions of the genomic HDV ribozyme. *Nucleic Acids Res.* **20**, 3919–3924 (1992).
- Kumar, P. K. R., Suh, Y.-A., Taira, K. & Nishikawa, S. Point and compensation mutations to evaluate essential stem structures of genomic HDV ribozyme. *FASEB J.* **7**, 124–129 (1993).
- Kumar, P. K. R., Taira, K. & Nishikawa, S. Chemical probing studies of variants of the genomic hepatitis delta virus ribozyme by primer extension analysis. *Biochemistry* **33**, 583–592 (1994).
- Belinsky, M. G., Britton, E. & Dinter-Gottlieb, G. Modification interference analysis of a self-cleaving RNA from hepatitis delta virus. *FASEB J.* **7**, 130–136 (1993).
- Wadkins, T. S. & Been, M. D. Core-associated non-duplex sequences distinguishing the genomic and antigenomic self-cleaving RNAs of hepatitis delta virus. *Nucleic Acids Res.* **25**, 4085–4092 (1997).
- Rosenstein, S. P. & Been, M. D. Evidence that genomic and antigenomic RNA self-cleaving elements from hepatitis delta virus have similar secondary structures. *Nucleic Acids Res.* **19**, 5409–5416 (1991).
- Huang, Z.-S., Ping, Y.-H. & Wu, H.-N. An AU at the first base pair of helix 3 elevates the catalytic activity of hepatitis delta virus ribozymes. *FEBS Lett.* **413**, 299–303 (1997).
- Jeoung, Y.-H., Kumar, P. K. R., Suh, Y.-A., Taira, K. & Nishikawa, S. Identification of phosphate oxygens that are important for self-cleavage activity of the HDV ribozyme by phosphorothioate substitution interference analysis. *Nucleic Acids Res.* **22**, 3722–3727 (1994).
- Bravo, C., Lescure, F., Laugaa, P., Fourrey, J.-L. & Favre, A. Folding of the HDV antigenomic ribozyme pseudoknot structure deduced from long-range photocrosslinks. *Nucleic Acids Res.* **24**, 1351–1359 (1996).
- Otwiński, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–776 (1994).
- Terwilliger, T. C. & Kim, S.-H. Generalized method of determining heavy-atom positions using the difference Patterson function. *Acta Crystallogr. A* **43**, 1–5 (1987).
- de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
- Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. *et al.* Crystallography and NMR system: a new software system for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Brünger, A. T. Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
- Laskowski, R. J., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–290 (1993).
- Pleij, C. W. A., Rietveld, K. & Bosch, L. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.* **13**, 1717–1731 (1985).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insight from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
- Carson, M. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).

**Acknowledgements.** We thank K. Nagai for U1A-RBD expression plasmids; C. Ogata and D. Cook for help at beamline X4A of the National Synchrotron Light Source (Brookhaven National Laboratory); the staff of the Cornell High Energy Synchrotron Source for support at beamlines A-1 and F-1; R. Batey, D. Battle, J. Kieft, A. Luptak and R. Rambo for help with synchrotron data collection and useful comments; P. Adams, M. Been, S. Bellon, C. Correll, R. Gaudet, D. Herschlag, J. Ippolito, P. Moore, L. Silvan, T. Steitz, S. Strobel, O. Uhlenbeck and D. Wilson for helpful discussions; and the staff of the Yale Center for Structural Biology for crystallographic and computational support. A.R.F. was a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. This work was supported in part by grants from the Jane Coffin Childs Memorial Fund for Medical Research, the Searle Scholars program, the NIH, and a Beckman Young Investigator award. J.A.D. is an assistant investigator of the Howard Hughes Medical Institute, a Searle scholar, and a Fellow of the David and Lucile Packard Foundation.

Correspondence and requests for materials should be addressed to J.A.D. (e-mail: doudna@csb.yale.edu). Atomic coordinates have been deposited at the Protein Data Bank under accession number 1drz, and are also available directly from the authors (www.csb.yale.edu/people/doudna).

# Transient clouds in Titan's lower atmosphere

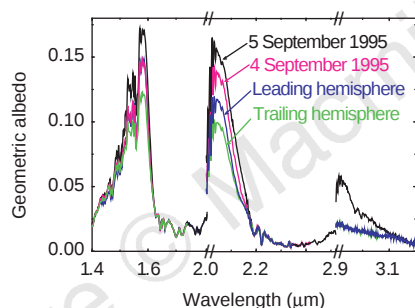
Caitlin A. Griffith\*, Tobias Owen†, Gary A Miller\* & Thomas Geballe‡

\* Department of Physics and Astronomy, Northern Arizona University, PO Box 6010, Flagstaff, Arizona 86001-6010, USA

† Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

‡ Joint Astronomy Centre, 660 N. Aohoku Place, Hilo, Hawaii 96720, USA

The 1980 encounter by the Voyager 1 spacecraft with Titan, Saturn's largest moon, revealed<sup>1,2</sup> the presence of a thick atmosphere containing nitrogen and methane (1.4 and ~0.05 bar, respectively). Methane was found to be nearly saturated at Titan's tropopause, which, with other considerations, led to the hypothesis that Titan might experience a methane analogue of Earth's vigorous hydrological cycle, with clouds, rain and seas<sup>3-7</sup>. Yet recent analyses of Voyager data indicate large areas of super-saturated methane, more indicative of dry and stagnant conditions<sup>8,9</sup>. A resolution to this apparent contradiction requires observations of Titan's lower atmosphere, which was hidden from the Voyager cameras by the photochemical haze (or smog) in Titan's stratosphere. Here we report near-infrared spectroscopic

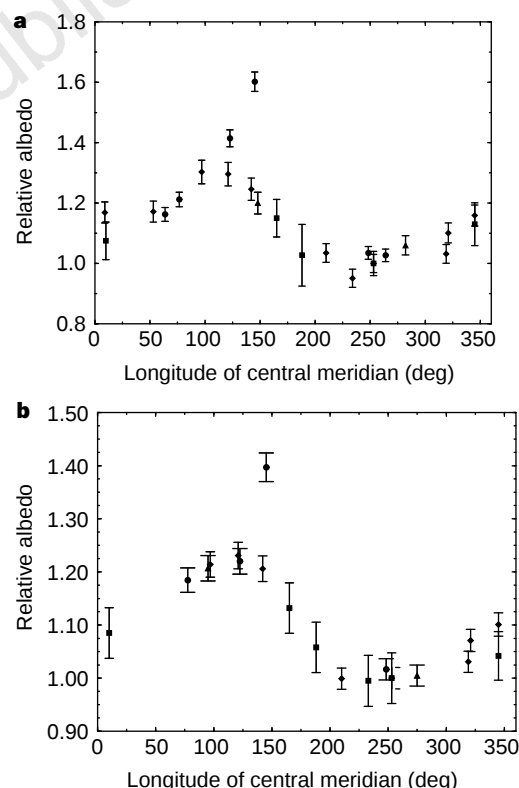


**Figure 1** UKIRT observations of Titan. Representative spectra of Titan's leading (blue line) and trailing (green line) hemispheres indicate the standard flux variations that result from the heterogeneity of Titan's surface. These data, recorded on 30 September 1993 (7:00 UT) and 7 October 1993 (9:00 UT), and ten nights of UKIRT observations agree with all previously published spectra. The dates, airmasses and the central meridians of all 12 nights of UKIRT observations are listed on our web site (<http://www.phy.nau.edu/~caitlin.griffith/titan.html>). The 2.9-μm spectra of Titan's leading and trailing hemispheres overlap as they are identical within the noise of the observations. Titan's exceptional spectra observed on 5 September 1995 at 12:00–13:00 UT (black line) display anomalously high reflectivities in the near-infrared windows at 1.3 (not shown), 1.6, 2.0 and 2.9 μm. The 4 September 1995 spectra, recorded at 10:00 UT (red line), display a smaller enhancement in albedo at 2.0 μm. Sections in spectral regions where the Earth's atmosphere is opaque are omitted. The 3σ error in the observations is 0.001 for the 1.2–2.3-μm spectra. The 2σ error in the 2.9–3.1-μm observations is 0.003. These spectra were recorded with two gratings and a variety of resolving powers from 350 to 1,200. Observations of Titan are interleaved with spectra of standard stars (BS8283, BS8418 and BS88), taken at airmasses that bracket that of Titan (airmass values range 1.03–1.40). Division of Titan's raw spectrum by that of the G stars eliminates telluric absorption features. Multiplication of this ratio by the blackbody flux of the G star provides a flux-calibrated spectrum of Titan. Division by a solar spectrum and correction factors for Titan's distances from Earth and Sun convert Titan's flux to a geometric albedo (the reflection from Titan relative to a Lambert surface, that is, diffuse perfect radiator). To allow comparison between observations (given photometric uncertainties), Titan's spectra are normalized to match in the low-albedo regions that sample Titan's stratosphere and, globally averaged, do not vary significantly<sup>10,12</sup>.

observations of Titan within four narrow spectral windows where the moon's atmosphere is ostensibly transparent. We detect pronounced flux enhancements that indicate the presence of reflective methane condensation clouds in the troposphere. These clouds occur at a relatively low altitude ( $15 \pm 10$  km), at low latitudes, and appear to cover ~9 per cent of Titan's disk.

Following the Voyager missions, ground-based observations established the visibility of Titan's lower atmosphere and surface at wavelengths of 0.9, 1.1, 1.3, 1.6, 2.0 and 5.0 μm<sup>10-13</sup>. During the past 10 years, spectra and images through these 'windows', and radar measurements, have revealed a heterogeneous surface underlying a uniform atmosphere devoid of thick variable clouds<sup>10-18</sup>.

Our observations, covering four spectral windows, were recorded over several years (1993–1997) with the United Kingdom Infrared Telescope (UKIRT), equipped with its spectrometer, CGS4. The 18 spectra, sampling many orientations of Titan's disk, replicate earlier measurements with two exceptions. On 5 September 1995, Titan's albedos exceeded those previously measured (at the same Titan longitude) by 14%, 17%, 30% and 200% at 1.3, 1.6, 2.0 and 2.9 μm, respectively. Observations taken a day earlier show a 14% increase in reflectivity at 2 μm, the most transparent of the two windows (1.6



**Figure 2** 2.0 μm (a) and 1.6 μm (b) light curves of Titan, derived from our observations (circles) and those recorded by Griffith *et al.*<sup>10</sup> (triangles), Coustenis *et al.*<sup>16</sup> (squares) and Lemmon *et al.*<sup>17</sup> (diamonds). This figure compares our data with observations taken over the past 10 years. Previous measurements document the reproducible variation in albedo, tied to Titan's rotational period<sup>10-18</sup>, and resulting from a heterogeneous surface. Indicating no other disturbance, they point to uniform atmospheric conditions devoid of thick variable clouds. At 2 μm, each point is the average albedo at 2.02–2.05 μm divided by the albedo average at 2.16–2.22 μm. At 1.6 μm, each point reflects an average albedo at 1.575–1.595 μm divided by the average at 1.69–1.74 μm. This division technique permits comparison between data sets independently of photometric uncertainties. Errors include the 3σ error due to random noise in Titan and the standard's spectra, as well as those due to incomplete elimination of telluric absorption. These values are scaled to 1 at 253° longitude, following the convention of ref. 18. The leading hemisphere faces the Earth at a sub-Earth longitude of ~87°; the trailing hemisphere is seen at ~268° longitude.

and 2  $\mu\text{m}$ ) recorded on that date (Figs 1 and 2). These exceptional observations are the focus of this report.

We present the 4 and 5 September 1995 observations in Fig. 1 along with examples of typical spectra of Titan's leading and trailing hemispheres. Like all previous measurements, they resemble a transmission spectrum of  $\text{CH}_4$ , and demonstrate how the large abundance of this gas ( $3 \times 10^{24}$ – $1.3 \times 10^{25}$  molecules/ $\text{cm}^2$ ) controls the opacity of Titan's atmosphere<sup>10</sup>. Low-albedo regions coincide with strong methane bands. Here, incident solar radiation is absorbed before reaching the troposphere<sup>10</sup>; the non-zero albedo results from sunlight scattered by the stratospheric haze. In contrast, the satellite's surface is seen through the haze at regions of weak  $\text{CH}_4$  absorption, corresponding to observed high albedos. Weak methane features within these windows exhibit constant depths and profiles, indicating that there is no appreciable variation in the methane abundance. Yet the average fluxes within these windows vary in a periodic fashion as a result of the heterogeneous reflectivity of Titan's surface<sup>10–18</sup> (Fig. 2).

The morphology of the enhanced fluxes on 4 and 5 September 1995—marked in Titan's windows but absent between windows—exonerates instrumental effects. Furthermore, spectra from standard stars obtained throughout every run agree with one another to within 2% at wavelength regions of good atmospheric transmission. The high quality of the UKIRT spectra indicates that the observed flux enhancements accurately document an unusual event in Titan's lower atmosphere.

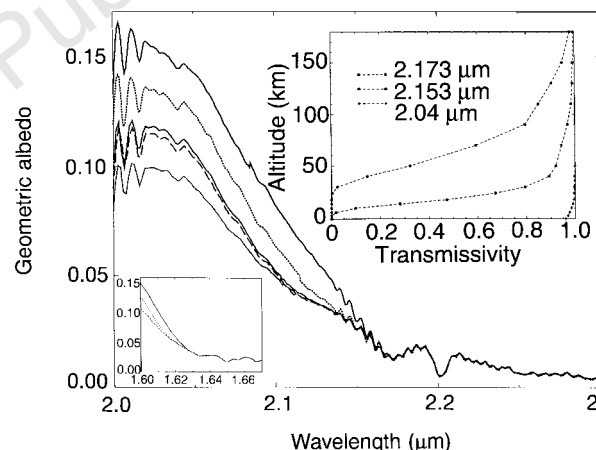
The cause of the abnormal window reflectivities of 4 and 5 September 1995 is addressed with radiative transfer calculations (described in ref. 10) that simulate absorption and scattering of incident solar radiation in Titan's atmosphere. Resultant synthetic spectra are compared with observations to derive properties of the atmosphere and surface. Titan's near-infrared spectrum results primarily from the absorption of sunlight by  $\text{CH}_4$ , the scattering of sunlight by haze, and the surface albedo. These effects can be isolated by studying separate segments of the spectrum. Applying this approach, we first analyse spectra of the undisturbed Titan to determine the optical depths of haze and  $\text{CH}_4$ , and derive surface reflectivities. After establishing a standard model, we address the peculiar 5 September 1995 spectrum, searching for a solution to explain the different magnitudes of enhancement in each window, their spectral ranges, and the small enhancement at 2.0  $\mu\text{m}$  on the previous day.

Spectral regions within the strong  $\text{CH}_4$  bands sample altitudes of 30–90 km. At these wavelengths,  $\text{CH}_4$  lines are saturated and Titan's

non-zero reflectivity, insensitive to uncertainties in the  $\text{CH}_4$  abundance, directly measures sunlight scattered off Titan's haze<sup>10</sup>. We assume spherical particles with the reflectivity properties of 'tholins', laboratory analogues of Titan's haze<sup>19</sup>. Because  $\text{CH}_4$  parameters are known<sup>20</sup> at 1.64 and 2.17  $\mu\text{m}$ , we focused on these spectral regions, and derived haze optical depths of  $0.5 \pm 0.1$  and  $0.20 \pm 0.04$ , respectively. These values indicate an effective particle size of  $0.4 \pm 0.2 \mu\text{m}$ , which is consistent with earlier studies<sup>21–23</sup>, and provide estimates of the haze optical depth within the spectral windows (Table 1).

Laboratory spectra indicate low  $\text{CH}_4$  optical depths ( $\tau < 0.05$ ) within the clearest regions of the 1.3, 1.6 and 2.0  $\mu\text{m}$  windows<sup>24</sup>. The higher value at 2.915  $\mu\text{m}$  ( $\tau = 0.4$ ) is transparent enough to allow visibility to the surface, confirming that 2.8  $\mu\text{m}$  is indeed a window.

UKIRT spectra reveal surface albedos for Titan's leading hemisphere resembling those of Jupiter's moon Callisto (Table 1). The surface albedo derived at 2.9  $\mu\text{m}$  was dark, as expected. However, unexpectedly it did not change from the leading to the trailing hemisphere. This contrasts with observations at radar and shorter infrared wavelengths, which show an irregular continent-sized surface feature. Our data thus suggest that this feature consists of material that is dark at 2.9  $\mu\text{m}$ : potentially water ice, an ingredient of Titan's bulk composition, or haze precipitate.



**Figure 3** The transmission of sunlight in Titan's atmosphere. Typical spectra of different exposures of Titan's disk (from leading to trailing hemispheres) deviate at  $\sim 2.12 \mu\text{m}$ , where Titan's surface is visible. The middle solid line, dashed line and lower solid line are nominal spectra of Titan taken at central longitudes of 76, 53 and 236, respectively. Spectra taken on 4 September 1995 (dotted line) and 5 September 1995 (top solid line) deviate from nominal observations at longer wavelengths (for example at 2.15  $\mu\text{m}$ ), where the troposphere is probed, and  $\text{CH}_4$  absorption prevents visibility of the surface. Bottom inset: 5 September 1995 spectrum (solid line) is compared with the leading and trailing hemisphere data at 1.6  $\mu\text{m}$ . Top inset: the transmissivity ( $e^{-\tau}$ ) as a function of altitude in Titan's atmosphere for three wavelengths, convolved with a gaussian filter with a full-width at half-maximum of  $3.8 \text{ cm}^{-1}$ . At 2.15  $\mu\text{m}$ , solar radiation is absorbed before reaching the surface. The main source of atmospheric opacity,  $\text{CH}_4$ , was modelled with a vertical distribution in which the humidity is specified at the surface and a constant mixing ratio is assumed from the surface up to altitudes at which the mixing ratio exceeds the saturated value. Here a saturation profile is assumed up to the tropopause. A constant mixing ratio (the saturated value) is assumed for the stratosphere. Although a range of humidities is considered, our nominal  $\text{CH}_4$  profile assumes 50% humidity at the surface, and a stratospheric mixing ratio of 0.02. The supersaturated profiles of ref. 8 are also consistent with our observations.  $\text{CH}_4$  line intensities and energies were derived by using line-by-line models from the raw laboratory data kindly provided by Strong *et al.*<sup>24</sup>. This technique allowed us to ascertain uncertainties in the continuum of the laboratory measurements (for example resulting from water ice absorption) that particularly affect regions of weak methane absorption characteristic of the windows of Titan's atmosphere.

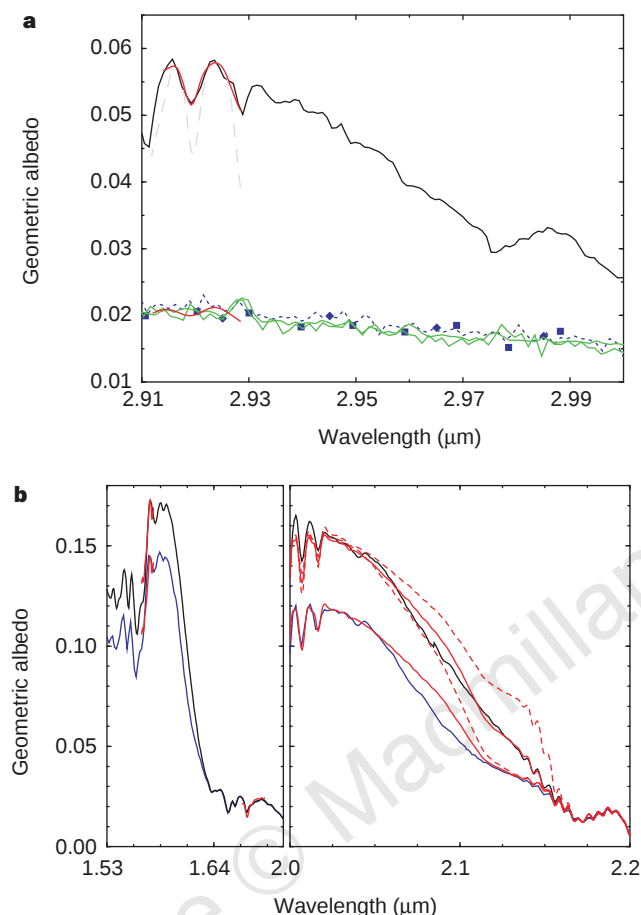
**Table 1** Surface albedos, derived haze optical depths and cloud coverage at various wavelengths

| Variable            | 1.28 $\mu\text{m}$ | 1.59 $\mu\text{m}$ | 2.0 $\mu\text{m}$ | 2.9 $\mu\text{m}$ |
|---------------------|--------------------|--------------------|-------------------|-------------------|
| Albedo              |                    |                    |                   |                   |
| Leading hemisphere  | 0.14               | 0.13               | 0.1               | 0.03              |
| Trailing hemisphere | —                  | 0.1                | 0.007             | 0.03              |
| Callisto            | 0.19               | 0.18               | 0.16              | 0.05              |
| Haze optical depth  | 0.8                | 0.52               | 0.22              | 0.1               |
| Cloud coverage (%)  | $9 \pm 1$          | $8 \pm 1$          | $9 \pm 1$         | $9 \pm 1$         |

The haze optical depths possess errors of 10% from uncertainties in the derived particle size and distribution. An uncertainty of 20% in the absolute albedo for Titan implies an additional error of 20% in the haze optical depth. The latter error affects the optical depth at different wavelengths similarly. We show the cloud coverages at 15 km altitude needed to enhance the window albedos to those observed on 5 September 1995. The value inferred from the 1.3  $\mu\text{m}$  window derives from the average of the 1.282–1.290  $\mu\text{m}$  flux ratio to that at 1.377–1.386  $\mu\text{m}$ . This ratio was observed to be 11.1 on 5 September 1995. The 1.28  $\mu\text{m}$  surface albedo for Titan's leading hemisphere was derived from data in ref. 17 and from NASA's Infrared Telescope Facility measurements taken on 25 July 1990. Other leading-hemisphere albedos come from UKIRT observations taken on 14 and 30 September 1993 and 26 October 1997. Trailing-hemisphere albedos at 1.59, 2.0 and 2.9  $\mu\text{m}$  were derived from data from 7 October 1993, 4 August 1993 and 27 and 28 August 1997. Comparison between data sets is, however, restricted to relative albedos, as a result of photometric errors. The albedos that we measured increase with decreasing wavelength, with the albedos at 1.6 and 2.0  $\mu\text{m}$  equal within errors. This trend was also derived (for the 1.3, 1.6 and 2.0  $\mu\text{m}$  windows) in refs 10 and 18. However, a different trend was determined in ref. 17. A spectrum<sup>25</sup> of Callisto's leading hemisphere resembles Titan's icy bedrock is exposed, even though haze particles and liquids ( $\text{C}_2\text{H}_6$ ) must precipitate and have been accumulating on Titan's surface since the formation of its atmosphere.



Having established a standard model, we address the anomalous 4 and 5 September 1995 spectra. The confinement of enhanced fluxes to Titan's windows indicates a perturbation of the lower atmosphere. Because the enhancements extend into spectral regions insensitive to Titan's surface, for example at 2.14–2.15 and 1.62–1.625  $\mu\text{m}$  (Fig. 3), Titan's surface is not the cause. Weak  $\text{CH}_4$



**Figure 4** Comparisons between radiative transfer models (red lines) and observations. **a**, UKIRT observations at 2.9  $\mu\text{m}$  taken on 5 September 1995 (black line) contrast with standard observations of Titan's leading hemisphere (observed on: 25 July 1990 (squares), 27 May 1987 (diamonds) and 26 October 1997 (dashed blue line)) and trailing hemisphere (observed on 27 and 28 August 1997 (green lines)). The 1987 and 1990 observations, having a spectral resolution of 150, were taken at NASA's Infrared Telescope Facility with the CGAS spectrometer. Unlike in other windows, at 2.9  $\mu\text{m}$   $\text{CH}_4$  signatures are affected by the presence of clouds, which act as reflecting beacons from which two-thirds of Titan's flux originates. The dashed grey line shows a calculation that assumes no clouds and a surface albedo adjusted to 0.12 to fit the overall flux. We consider the range of possible  $\text{CH}_4$  distributions (25% humidity to supersaturation); even the lowest value, with the shallowest features (25% shown), produces features deeper than those observed. In contrast, an optically thick cloud at 15 km altitude that covers 9% of the planet reproduces the features (red line). Here, the surface albedo was established by modelling Titan's standard spectrum (blue and green lines), assuming cloud-free conditions (lower red line). **b**, We interpret spectra of Titan's leading hemisphere (blue lines) with a cloud-free model (bottom red lines). The 5 September 1995 spectrum (black line), analysed with the same surface reflectivities, haze and  $\text{CH}_4$  optical depths, requires a thick cloud, covering 9% of the surface. The three models shown (red lines at 2  $\mu\text{m}$ ) differ only in the altitude level at which the cloud was placed. Clouds at 4 and 24 km altitude (top and bottom dashed lines) fail to interpret the spectra, whereas a 15-km cloud (solid red line at 1.6 and 2  $\mu\text{m}$ ) fits well. Omissions in spectral-line information probably account for the model's deviation from observations at 2.08  $\mu\text{m}$ . Terrestrial  $\text{CO}_2$  degrades the spectrum at 2.0–2.2  $\mu\text{m}$  as a result of incomplete elimination of telluric absorption.

signatures at 2.9  $\mu\text{m}$  also eliminate the possibility that surface variations produced the elevated fluxes; the calculated strengths of the 2.9  $\mu\text{m}$  features exceeded those observed when interpreted with an increase in the surface albedo (Fig. 4a).

To investigate the effects of condensation, we add highly reflective ( $\omega = 0.98$ ) and optically thick clouds to our standard model of Titan. We assume a uniform cloud optical depth throughout the 1.3–3.1  $\mu\text{m}$  region, a behaviour expected for  $\text{CH}_4$  particles exceeding 5  $\mu\text{m}$  in size. Large particle sizes and rapid particle growth are predicted in the event of  $\text{CH}_4$  condensation on Titan, because condensation nuclei are thought to be scarce<sup>5</sup>. A consistent explanation for all the observations emerges from this analysis. The spectral characteristics of the flux enhancements in each window independently point to the same cause: a thick cloud deck located at 15 km altitude and covering 7–9% of Titan's disk.

Several lines of evidence support this interpretation. Clouds at 15 km altitude enhance Titan's flux at precisely the spectral ranges observed (Fig. 3). Deeper clouds affect a narrower spectral range for each window; higher clouds would be observed over greater spectral ranges (Fig. 4b). Second, the same ~7–9% coverage of clouds located at 15 km reproduces the observed flux enhancements at all windows (Fig. 4). In contrast, low clouds (for example, at 5 km altitude) require 9% covering at 1.28, 1.59 and 2.0  $\mu\text{m}$ , and 15% at 2.9  $\mu\text{m}$ . This is a consequence of the increased  $\text{CH}_4$  absorption at 2.9  $\mu\text{m}$ . Last, a cloud deck at 15 km altitude with 9% coverage interprets the 2.9  $\mu\text{m}$  spectrum well; deeper (higher) clouds indicate features stronger (weaker) than observed.

On 4 September 1995, observations show a 14% enhancement in the reflectivity at 2.0  $\mu\text{m}$  (Fig. 2). No enhancement occurred at 1.6  $\mu\text{m}$ . On 5 September, the 17% enhancement at 1.6  $\mu\text{m}$  was dwarfed by the 30% change at 2  $\mu\text{m}$ . These observations can be naturally interpreted with the presence of low-latitude clouds on Titan's limb on 4 September (i.e. centred near the equator having an airmass of 2) that rotated into better view the following day. Unfortunately, no observations were scheduled on the day after 9 May 1995 to allow further investigation. When follow-up observations were recorded in 1997, all was normal.

At present it is too early to address the cause for a sudden appearance of clouds on Titan. Cooler and more massive, Titan's atmosphere is not expected to be as turbulent as Earth's. Yet some vertical motion must accompany the meridional motions that equalize the surface temperatures. In addition, clouds from Titan have been hypothesized to be brief events and therefore difficult to detect. Titan's haze, the only seed nuclei identified so far, has a low number density<sup>5,8</sup>. Thus, in comparison with Earth, particle growth in clouds might be rapid, and closely followed by precipitation and cloud dissipation.

Whether clouds and rain are occasional, although quick, occurrences that punctuate otherwise dry conditions can be addressed by future observations. If caused by convection, cloud formation will be revealed with additional near-infrared data indicating the mass, altitude and duration of the clouds. If volcanic or orographic, the clouds will correlate with Titan's longitude. Titan presents us with a unique opportunity to understand processes native to our biosphere, yet in a cooler atmosphere ten times more massive than Earth's. □

Received 19 March; accepted 8 July 1998.

1. Lindal, G. F. *et al.* The atmosphere of Titan: an analysis of the Voyager 1 radio-occultation measurements. *Icarus* **53**, 348–363 (1983).
2. Lellouch, E. *et al.* Titan's atmospheric temperature profile: a reanalysis of the Voyager 1 radio-occultation and IRIS 7.7 micron data. *Icarus* **79**, 328–349 (1989).
3. Lunine, J. I., Stevenson, D. J. & Yung, Y. L. Ethane ocean on Titan. *Science* **222**, 1229–1230 (1983).
4. Flasar, F. M. Oceans on Titan? *Science* **221**, 55–57 (1983).
5. Toon, O. B., McKay, C. P., Courtin, R. & Ackerman, T. P. Methane rain on Titan. *Icarus* **75**, 255–284 (1988).
6. Eschleman, V. R., Lindal, G. F. & Tyler, G. L. Is Titan wet or dry? *Science* **221**, 55–57 (1983).
7. Hunten, D. M. *et al.* in *Saturn* (eds Gehrels, T. & Matthews, M. S.) 671–759 (Univ. of Arizona Press, Tucson, 1984).
8. Courtin, R., Gautier, D. & McKay, C. Titan's thermal emission spectrum: reanalysis of the Voyager infrared measurements. *Icarus* **114**, 144–162 (1995).

9. Samuelson, R. E., Nath, N. R. & Borysow, A. Gaseous abundances and methane supersaturation in Titan's tropopause. *Planet. Space Sci.* **45**, 959–980 (1997).
10. Griffith, C. A., Owen, T. & Wagoner, R. Titan's surface and troposphere, investigated with ground-based, near-infrared observations. *Icarus* **93**, 362–378 (1991).
11. Lemmon, M. T., Karkoschka, E. & Tomasko, M. Titan's rotation: surface feature observed. *Icarus* **103**, 329–332 (1993).
12. Griffith, C. A. Evidence for surface heterogeneity on Titan. *Nature* **364**, 511–513 (1993).
13. Noll, K. S., Geballe, T. S., Knacke, R. F. & Pendleton, Y. J. Titan's 5  $\mu\text{m}$  spectral window: carbon monoxide and the albedo of the surface. *Icarus* **124**, 625–631 (1996).
14. Muhleman, D. O., Grossman, A. W., Butler, B. J. & Slade, M. A. Radar reflectivity of Titan. *Science* **248**, 975–980 (1990).
15. Smith, P. H. et al. Titan's surface revealed by HST imaging. *Icarus* **119**, 336–349 (1996).
16. Combes, M. et al. Spatially resolved images of Titan by means of adaptive optics. *Icarus* **129**, 482–497 (1997).
17. Lemmon, M. T., Karkoschka, E. & Tomasko, M. Titan's rotational light-curve. *Icarus* **113**, 27–38 (1995).
18. Coustenis, A., Lellouch, E., Maillard, J. P. & McKay, C. P. Titan's surface: composition and variability from its near-infrared albedo. *Icarus* **118**, 87–104 (1995).
19. Khare, B. N. et al. Optical constants of organic tholins produced in a simulated Titanian atmosphere: from soft X-ray to microwave frequencies. *Icarus* **60**, 127–137 (1984).
20. Hesson, N., Bonnet, B., Scott, N. A. & Chédin, A. The "GEISA" data bank, 1991 version (Int. note L.M.D. 163, Lab. de Météorologie Dynamique, Palaiseau, 1991).
21. Rages, K., Pollack, J. B. & Smith, P. H. Size estimates of Titan's aerosols based on Voyager high-phase-angle images. *J. Geophys. Res.* **88**, 8721–8728 (1983).
22. McKay, C. P., Pollack, J. B. & Courtin, R. The thermal structure of Titan's atmosphere. *Icarus* **80**, 23–53 (1989).
23. Toon, O. B., McKay, C. P., Griffith, C. A. & Turco, R. P. A physical model of Titan's aerosols. *Icarus* **95**, 24–53 (1992).
24. Strong, K., Taylor, F. W., Calcutt, S. B., Remedios, J. J. & Ballard, J. Spectral parameters of self- and hydrogen-broadened methane from 2000 to 9500  $\text{cm}^{-1}$  for remote sounding of the atmosphere of Jupiter. *J. Quant. Spectrosc. Rad. Transfer* **50**, 363–429 (1993).
25. Calvin, W. M. & Clark, R. N. Modeling the reflectance spectrum of Callisto 0.25 to 4.1  $\mu\text{m}$ . *Icarus* **89**, 305–317 (1991).

**Acknowledgements.** We thank C. McKay for conversations on the weather, M. Flasar for suggestions that improved the paper, K. Noll for a helpful reading and C. B. Phillips for help in operating arduous radiative transfer codes. C.G. was supported by an NSF Young Investigator Award, the Research Corporation, and the NASA Planetary Astronomy Program.

Correspondence and requests for materials should be addressed to C.A.G. (e-mail: griffith@cna.nau.edu).

## Mesoscopic behaviour of the neutral Fermi gas $^3\text{He}$ confined in quantum wires

S. L. Phillipson\*, A. M. Guénault\*, S. N. Fisher\*, G. R. Pickett\* & P. J. Y. Thibault†

\* Department of Physics, Lancaster University, Lancaster LA1 4YB, UK

The behaviour of electron gases in restricted geometries provides a means to explore the fundamental quantum-mechanical properties of fermion gases at mesoscopic length scales<sup>1</sup>. But the existence of Coulomb repulsion between electrons unavoidably complicates the physics. Quantum gases of neutral fermions—such as  $^3\text{He}$  quasiparticles in a dilute solution of  $^3\text{He}$  in  $^4\text{He}$ , cooled to millikelvin temperatures<sup>2</sup>—therefore offer a means of probing regimes completely inaccessible to electronic systems. Here we demonstrate the quantum exclusion of a  $^3\text{He}$  fermion gas from a network of narrow channels, connected to a reservoir of  $^3\text{He}/^4\text{He}$  solution. The effect is expected from simple quantum-mechanical arguments, which predict that the  $^3\text{He}$  atoms cannot enter the channels when their wavelength exceeds  $\sqrt{2}$  times the channel width. By adjusting the temperature of the solution, the energy of the particles and hence their average wavelength can be controlled. In this way, we observe temperature-dependent changes in the penetration of the  $^3\text{He}$  quasiparticles into the channels. Our results demonstrate the macroscopic response of an atomic gas to basic quantum-mechanical restrictions at the mesoscopic level.

Although the  $^3\text{He}$  gas in dilute  $^3\text{He}/^4\text{He}$  solutions might be expected to show broadly similar behaviour to that of an electron

gas when similarly confined, this Fermi–Dirac quantum fluid has a number of advantages over the usual metallic or semiconducting systems. The medium can be readily decanted into a container shaped to provide the requisite restricted geometry and therefore various solutions can be studied under identical conditions. The  $^3\text{He}$  solutions are absolutely clean at millikelvin temperatures and thus there are no impurity effects. The Fermi parameters of the  $^3\text{He}$ —Fermi energy  $E_F$ , Fermi temperature  $T_F$  and Fermi wavelength  $\lambda_F$ —can be varied over a wide range by adjustment of the  $^3\text{He}$  concentration. Finally, unlike in the charged electron gas, the potential energy determining the bottom of the excitation band is only a very weak function of the local particle density and there are no problems of Coulomb repulsion.

The Fermi temperature of a dilute  $^3\text{He}/^4\text{He}$  solution of fractional  $^3\text{He}$  concentration  $x$  varies as  $x^{2/3}$ . A 0.1%  $^3\text{He}$  solution yields a Fermi energy of around 26 mK with an associated Fermi wavelength of 8 nm. This is much smaller than would be practicable in electronic systems, but with helium solutions we can make use of confined geometries on this scale over the temperature range 5–100 mK which is readily accessible with a dilution refrigerator.

The experiment reported here, the detection of the quantum exclusion of the  $^3\text{He}$  gas from a system of narrow channels as the temperature is reduced, was chosen for two reasons. First, it is conceptually simple, and second, it has no direct analogue in an electronic system, as quantum exclusion of an electron gas leads to a change in the local potential which largely cancels the effect. Furthermore, a metallic system cannot be probed over a temperature range spanning  $T_F$ .

The principle of an ideal experiment is illustrated in Fig. 1. A bulk  $^3\text{He}/^4\text{He}$  solution is placed in contact with a network of narrow cavities with a scale dimension of  $a$ . The  $^3\text{He}$  concentration is adjusted such that the Fermi wavelength is longer than  $\sqrt{2}a$ , or in other words, the confinement energy ( $E_0$ , of the order of  $\hbar^2/4m^*a^2$ , where  $m^*$  is the  $^3\text{He}$  effective mass) is greater than  $kT_F$ . Therefore, at temperatures well below  $T_F$ ,  $^3\text{He}$  atoms cannot penetrate into the cavities. However, as the temperature is increased, at some point the most energetic atoms begin to penetrate. This starts to occur significantly when  $(E_F + 2kT)$  exceeds the confinement energy  $E_0$ . Therefore, if over the temperature range used the quantity  $(E_F + 2kT)$  can be made to span the confinement energy, then at the low-temperature end  $^3\text{He}$  is totally excluded from the pores but will be able to penetrate into them at the higher temperatures. We should therefore see an increase in the concentration of  $^3\text{He}$  in the pores with increasing temperature.

The confined geometry is provided by the channels in porous Vycor glass which have<sup>3</sup> a characteristic dimension  $a$  of around 7 nm. Given the accepted value<sup>2</sup> for the effective mass  $m^*$  for  $^3\text{He}$  quasiparticles as 2.25 times the bare  $^3\text{He}$  atomic mass, the expected confinement energy  $E_0$  is 15 mK in temperature units. This is a little low to implement the ideal experiment, as the lowest reliable temperature to which we can cool the cell is around 8 mK, meaning that  $kT$  is rather large even at our base temperature. However, calculations show that even with a 0.1%  $^3\text{He}$  solution there should be a significant change of the pore penetration with temperature, and this concentration was chosen. For more dilute solutions much of the variation would occur below 8 mK, and also the measurable effect would be smaller.

We infer the extent of the penetration of  $^3\text{He}$  into the pores from the concentration of  $^3\text{He}$  in the external bulk solution; the bulk concentration is monitored by a capacitive measurement of the dielectric constant of the solution. As the changes in dielectric constant with changes in  $^3\text{He}$  concentration are small, we must ensure that the volume of bulk solution in our experimental cell is as small as possible in comparison with the pore volume. The bulk volume must accommodate both the capacitor for measuring the  $^3\text{He}$  concentration and pads of silver sinter necessary to make thermal contact to the solution. (The sinter pore size is  $\sim 100$  nm

† Present address: Centre de Recherche sur les Très Basses Températures, CNRS, BP166, 38042 Grenoble cedex 9, France.

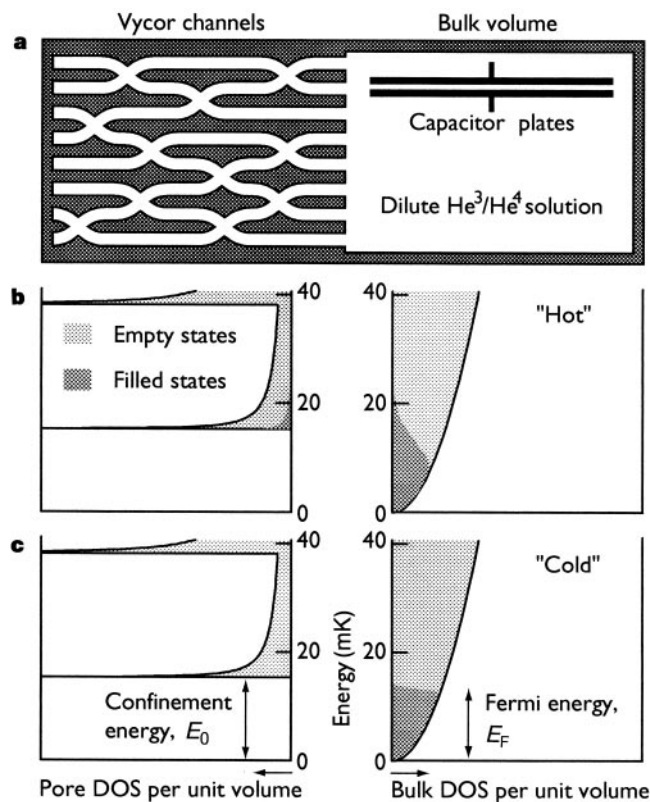
and hence qualifies as "bulk" volume.) The pore volume in the Vycor represents 17% of the total helium volume of the cell. This figure is determined from the quantity of helium required to fill the cell along with the known volume of the Vycor glass, and the knowledge<sup>3</sup> that the pores in the Vycor represent 30% of the bulk Vycor volume. The cell is thermally anchored to the mixing chamber of a 2 mK dilution refrigerator<sup>4</sup>.

The principle of the measurement is straightforward. The dilute solution fills the adjacent bulk and confined regions. At low temperatures, the energy of a significant fraction of the  $^3\text{He}$  atoms in the solution is lower than the pore confinement energy, and these atoms are excluded from the Vycor. The  $^4\text{He}$  component being superfluid can penetrate freely. As the temperature is increased, the

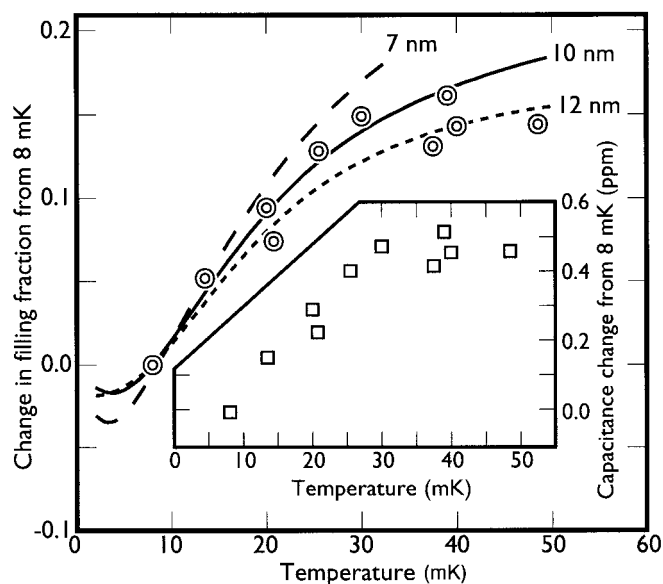
Fermi distribution becomes wider, the fraction of  $^3\text{He}$  atoms excluded falls, and thus the degree of penetration into the pores increases. This leads to a fall in the  $^3\text{He}$  concentration in the bulk region which we can monitor. Since the  $^3\text{He}$  atomic volume is larger than that of  $^4\text{He}$ , with falling  $^3\text{He}$  concentration the number of helium atoms in the capacitor rises, hence increasing the capacitance. Thus we should observe the capacitance to increase with increasing temperature.

To perform the experiment, we monitor the capacitance as a function of temperature to infer the changes in  $^3\text{He}$  density in the Vycor. This procedure requires much care to avoid temperature gradients which can drive  $^3\text{He}$  around the cell. We take 8 mK as the baseline, measure the capacitance and then monitor the change in capacitance as the temperature is increased. The temperature of the cell is adjusted by stabilizing the temperature of the mixing chamber of the dilution refrigerator, achieved by heating the final heat exchanger. Thermal equilibrium in the cell is typically reached only some 3 hours after the temperature is changed; thereafter, the cell must be left for at least a further 3 hours to establish the appropriate capacitance value. The heating is then stopped and after a further similar period the 8 mK background value checked. The practical range is limited by the lowest temperature at which we can reliably maintain the cell in thermal equilibrium with the mixing chamber, 8 mK, and the upper limit of 50 mK above which measurements become difficult owing to increasing temperature instabilities.

As the capacitance changes are so small, we first have to establish the background variation of the capacitance with temperature. We can either measure the capacitance of the  $^3\text{He}/^4\text{He}$  solution as a function of temperature with no Vycor in the cell (that is, when there is no effect on the  $^3\text{He}$  concentration from any quantum confinement), or we can measure the capacitance changes for pure  $^4\text{He}$  with the Vycor present. Both methods yield the same result,



**Figure 1** Schematic view of the experiment. In **a**, a dilute  $^3\text{He}/^4\text{He}$  solution is contained in a bulk volume in contact with a set of interconnecting 'quantum wires', which in the present case comprise the pores in Vycor glass. Ideally the Fermi energy of the solution should be adjusted to be below the confinement energy of the pores. **b** and **c**, schematic views of the densities of states (DOS) and the disposition of the  $^3\text{He}$  component in the solution between the two regions as a function of temperature. In **b** the temperature is 'high', and  $^3\text{He}$  atoms can penetrate into the confined wires, whereas in **c**, where the temperature  $T \ll T_F$ ,  $^3\text{He}$  is excluded from the restricted volume as there are no occupied states above the confinement energy  $E_0$ . The DOS in the quantum wire region is calculated on the basis of the uniform tube model discussed in the text. As the pores are randomly connected, the DOS for the pores would be qualitatively similar but without the large discontinuities shown here. The 'bulk' volume contains a capacitor to measure the  $^3\text{He}$  concentration. This is a formidable task because the difference in dielectric constant between  $^3\text{He}$  and  $^4\text{He}$  is rather small. (The dielectric constant of the solution<sup>6,7</sup> is  $\epsilon = 1.0572 - 0.0166x$  where  $x$  is the fractional  $^3\text{He}$  concentration.) The expected change in capacitance is of the order of 1 part in  $10^6$ , which required us to develop a capacitor with at least 1 part in  $10^7$  stability for satisfactory measurements. The  $\sim 10$  pF capacitance is measured with an Andeen-Hagerling model 2500 1 kHz automatic capacitance bridge. We monitor the temperature of the cell with a vibrating wire resonator (as described in ref. 4) in the mixing chamber of the dilution refrigerator.



**Figure 2** The measured penetration of the  $^3\text{He}$ . Inset, the change in relative capacitance with temperature of 0.1%  $^3\text{He}/^4\text{He}$  solution as a result of the quantum penetration of  $^3\text{He}$  atoms into the Vycor with increasing temperature. Main figure, the same data converted to the change in relative filling factor in the pores. The curves represent the change in filling factor calculated from a simple uniform 'quantum wire' model as described in the text for pore channel widths of 7, 10 and 12 nm. The absolute filling factor changes from about 0.4 to 0.6 for the 10 nm channels over this temperature range. (The upturn in the calculations below 5 mK arises from the incomplete exclusion at  $T = 0$  together with a DOS which falls with increasing energy, see Fig. 1).



namely a fall of about 1.0 p.p.m. in capacitance over the temperature range used. The agreement between the two sets of data, taken in separate runs a few months apart, shows that the capacitor behaves reproducibly at a level of one in  $10^7$  over a long time period.

Knowing the background capacitance, we can investigate the effect of the quantum exclusion. The measured change in capacitance for a 0.1%  $^3\text{He}$  in  $^4\text{He}$  solution with the background subtracted is shown in Fig. 2 inset. The main graph in Fig. 2 shows the data converted to  $^3\text{He}$  filling factor in the pores, defined as the ratio of  $^3\text{He}$  concentration in the pores to total  $^3\text{He}$  concentration in the solution—that is, complete penetration of  $^3\text{He}$  into the pores corresponds to a filling factor of unity. The conversion requires no adjustable parameters as we know all the volumes in the cell. Figure 2 shows that the filling factor in the pores has increased by 0.2 over our temperature range. The effect of the quantum exclusion with the gradual expulsion of  $^3\text{He}$  from the confined region as the temperature falls is quite clear. Had we been able to work at a lower temperature and with a concentration approaching the ideal situation of Fig. 1, we should have been able to span a temperature range where the filling factor changed from near zero to unity—that is, from complete exclusion to complete penetration.

We can compare the observed  $^3\text{He}$  penetration with a calculation made on the basis of a simple single-particle model. Although Vycor glass has a complicated structure, for a zeroth-order model we assume the pores to constitute an array of uniform quantum wires with a square cross-section of width  $a$  and with the known total volume. The  $^3\text{He}$  density of states is then given by the combination of transverse standing-wave states in the square cross-section and the free longitudinal motion. This gives a density of states as a function of energy  $\epsilon$  made up of a number of  $(\epsilon - \epsilon_0)^{-1/2}$  terms in which  $\epsilon_0 = (n^2 + l^2)E_0/2$  with  $E_0 = \hbar^2/4m^*a^2$ , and where  $n$  and  $l$  are the positive integer quantum numbers defining the transverse state. (The two lowest such terms are those shown in Fig. 1.)

This calculation has just one adjustable parameter, the confinement energy  $E_0$ , or equivalently the channel width  $a$ . We note that  $a$  here is the width open to  $^3\text{He}$  quasiparticles; this is the bare Vycor channel width minus  $\sim 1$  nm to allow for the two atomic layers of solid  $^4\text{He}$  adsorbed on the wall. The density of states so derived may be directly compared with the experiment because there are no other unknowns. We assume that the states are filled with the Fermi Dirac distribution, and compute the chemical potential needed to ensure that the correct total number of  $^3\text{He}$  quasiparticles occupies the cell. The three curves in Fig. 2 show the filling factor calculated in this way on the assumption that  $a = 7, 10, 12$  nm, that is, that  $E_0 = 5, 7.5, 15$  mK. Given the simplicity of the assumptions, the agreement between the observed filling factor and the quantum exclusion calculation for  $a \approx 10$  nm is good over the whole temperature range.

This agreement confirms that there is indeed a change in the quantum confinement of  $^3\text{He}$  quasiparticles in narrow channels as a function of temperature. This is, to our knowledge, the first confinement experiment to be performed in helium; with more refined techniques we expect to gain access to a whole range of mesoscopic behaviour which has no analogue in electronic systems. For example, the equivalent of the metal/insulator transition near the critical point is complicated in the charged system by the Coulombic repulsion between the electrons. It would also be interesting to study more dilute solutions and/or narrower channels where  $E_F$  could be adjusted to be much smaller than the confinement energy, as with low enough temperatures we would see the complete transition from exclusion to penetration. Finally, we could add an energy  $\sim \pm 10$  mK to the confinement energy by applying a high magnetic field, making the quantum exclusion from the channels dependent on the quasiparticle spin. This would greatly enhance the effect for the antiparallel spin, and suggests that a porous membrane in a high magnetic field might provide a  $^3\text{He}$  polarization filter<sup>2</sup>. □

Received 28 May; accepted 27 July 1998.

1. Sohn, L. L., Kouwenhoven, L. P. & Schön, G. (eds) *Mesoscopic Electron Transport* (NATO ASI Ser, Vol. 345, Kluwer Academic, 1997).
2. Owers-Bradley, J. R. Spin-polarized  $^3\text{He}$ - $^4\text{He}$  liquids. *Rep. Prog. Phys.* **60**, 1173–1216 (1998).
3. Levitz, P., Ehret, G., Sinha, S. K. & Drake, J. M. Porous vycor glass—the microstructure as probed by electron microscopy, direct energy transfer, small angle scattering, and molecular adsorption. *J. Chem. Phys.* **95**, 6151–6161 (1991).
4. Bunkov, Yu. M. et al. A compact dilution refrigerator with vertical heat exchangers for operation to 2 mK. *J. Low Temp. Phys.* **83**, 257–272 (1991).
5. Ohi, A. & Nakayama, T. Percolative diffusion of dissolved  $^3\text{He}$  atoms in He II through porous media. *J. Low Temp. Phys.* **81**, 349–368 (1990).
6. Kierstead, H. A. Dielectric constant, molar volume, and phase diagram of saturated  $^3\text{He}$ - $^4\text{He}$  mixtures. *J. Low Temp. Phys.* **24**, 497–512 (1976).
7. Castelijn, C. A. M., Kuerten, J. G. M., de Waele, A. T. A. M. & Gijsman, H. M.  $^3\text{He}$  flow in dilute  $^3\text{He}$ - $^4\text{He}$  mixtures at temperatures between 10 and 150 mK. *Phys. Rev. B* **52**, 2870–2886 (1995).

**Acknowledgements.** We thank V. Fal'ko and A. Volkov for discussions.

Correspondence and requests for materials should be addressed to A.M.G. (e-mail: a.guenault@lancaster.ac.uk).

## Spin fluctuations in $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$

H. A. Mook\*, Pengcheng Dai\*, S. M. Hayden†, G. Aeppli‡, T. G. Perring§ & F. Doğan||

\* Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6393, USA

† H. H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, UK

‡ NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540, USA

§ ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK

|| Department of Materials Sciences and Engineering, University of Washington, Seattle, Washington 98195, USA

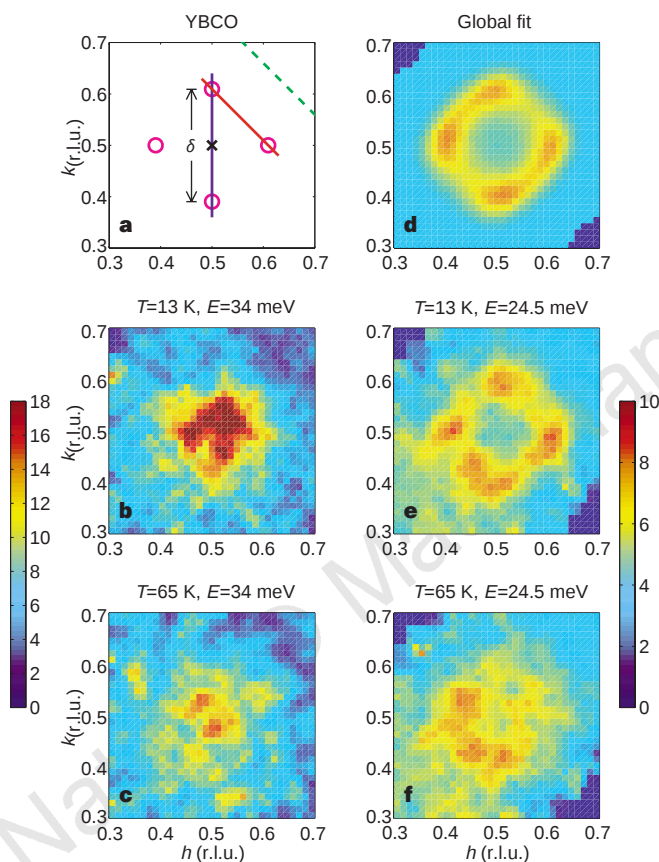
**An important feature of the high-transition-temperature ( $T_c$ ) copper oxide superconductors is the magnetism that results from the spins associated with the incomplete outer electronic shells ( $3d^9$ ) of the copper ions. Fluctuations of these spins give rise to magnetic excitations of the material, and might mediate the electron pairing that leads to superconductivity. If the mechanism for high- $T_c$  superconductivity is the same for all copper oxide systems, their spin fluctuations should be universal. But so far, the opposite has seemed to be the case: neutron scattering data reveal clear differences between the spin fluctuations for two major classes of high- $T_c$  materials,  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$  (refs 1–3) and  $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$  (refs 4–6), whose respective building blocks are  $\text{CuO}_2$  layers and bilayers. Here we report two-dimensional neutron-scattering imaging of  $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ , which reveals that the low-frequency magnetic excitations are virtually identical to those of similarly doped  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ . Thus, the high-temperature ( $T_c \lesssim 92$  K) superconductivity of the former materials may be related to spatially coherent low-frequency spin excitations that were previously thought to be unique to the lower- $T_c$  (<40 K) single-layer  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$  family.**

The most celebrated common feature of copper oxide magnetism is that the undoped parent compounds are insulating antiferromagnets, characterized by a simple doubling of the crystallographic unit cells in the  $\text{CuO}_2$  planes. Neutron scattering, which measures the Fourier transform (in space and time) of the spin-spin correlation function, long ago imaged the doubling<sup>7,8</sup> that manifests itself in diffraction peaks at the wavevectors  $(1/2, 1/2)$ . To label wavevectors in two-dimensional reciprocal space (Fig. 1a), we use reciprocal lattice units such that  $(1,0)$  and  $(0,1)$  are along the nearest-neighbour copper–oxygen–copper paths and are the locations of the lowest-order diffraction peaks from the nuclei in the nearly square  $\text{CuO}_2$  planes. Upon chemical doping to induce metallic and superconducting behaviours, the elastic diffraction peaks, due to static magnetic order, disappear. Both antiferromagnetic and superconducting compositions show strong inelastic scattering derived from magnetic fluctuations. For superconducting

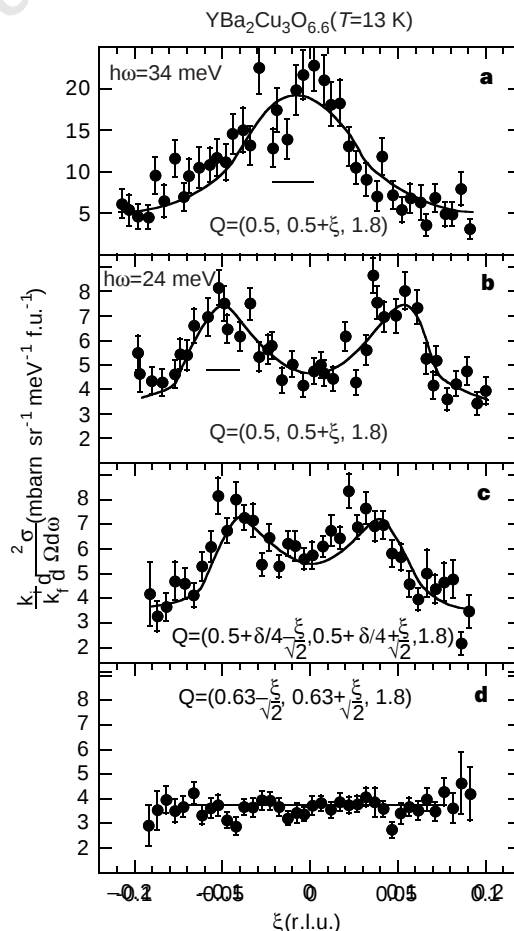
$\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$  (214), the low-frequency inelastic scattering peaks are not at  $(1/2, 1/2)$ , but at a quartet of nearby incommensurate wavevectors (indicated in Fig. 1a). As a function of frequency, no particularly sharp features have been identified. Chemical doping can be achieved by substitution of Sr, Ba or Ca for La or by intercalation of excess oxygen. The details of the scattering depend only on the unbound charge introduced by the dopants, and not on the method for introducing the charge<sup>9</sup>. For  $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$  (123), which so far has only been doped by excess oxygen, the inelastic scattering was thought to be dominated by broader maxima still centred at  $(1/2, 1/2)$ . Furthermore, when frequency is scanned, the scattering displays a sharp resonance whose position varies with  $T_c$

(refs 10, 11). Dai *et al.* have recently shown that for  $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$  ( $T_c = 62.7$  K), the situation is more complex<sup>12</sup>. Even though there is a sharp resonance centred at 34 meV and peaking at  $(1/2, 1/2)$ , the lower-frequency spin fluctuations actually peak away from  $(1/2, 1/2)$ . Because of experimental limitations, Dai *et al.*<sup>12</sup> were unable to determine either the precise displacements of the peaks from  $(1/2, 1/2)$  or the absolute amplitudes characterizing the fluctuations. A breakthrough in experimental technology, namely the two-dimensional position-sensitive detection of scattered neutrons in the High Energy Transfer time-of-flight neutron spectrometer at ISIS, the most powerful pulsed spallation neutron source, has now allowed us to image the fluctuations in the same 25.6-g sample. The outcome is a quantitative resolution of the issues left unresolved by Dai *et al.*<sup>12</sup> which permits direct comparison with (214) materials.

To our knowledge, these data are the first maps of the magnetic scattering ever produced in the two-dimensional reciprocal space for the  $\text{CuO}_2$  planes in (123). As the experiments are time-consuming (Fig. 1 legend), we chose two measuring frequencies. The first is  $34 \pm 2$  meV, where the resonance for this particular doping level occurs. Figure 1b and c illustrates the outcome at 13 K and 65 K ( $T_c + 2.3$  K). In agreement with previous measurements<sup>10,11</sup>, which were one-dimensional cuts through this map, there is a broad peak



**Figure 1** Reciprocal lattice diagram and neutron-scattering results. **a**, Two-dimensional reciprocal space for the  $\text{CuO}_2$  planes in the cuprates. The cross marks the  $(1/2, 1/2)$  point at which magnetic Bragg scattering appears in the parent insulating compounds; the quartet of circles indicates the locus of magnetic intensity maxima in the superconductors measured in reciprocal lattice units (r.l.u.). **b**, Image of the 34 meV resonance at 13 K for  $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ . The crystal and neutron-scattering instrument are three-dimensional, so that we are observing a two-dimensional cut through reciprocal space, with the out-of-plane wavevector component integrated from  $(\pi/d)(1 - 0.5)$  to  $(\pi/d)(1 + 0.5)$  where  $d = 3.342$  Å, the interlayer spacing. The colour bar gives intensities in the absolute units of  $\text{mbarn sr}^{-1} \text{meV}^{-1} \text{F.U.}^{-1}$ . **c**, Image at the resonance frequency on warming to 65 K. **e, f**, Images of the subresonance (24 meV) scattering at 13 and 65 K, respectively; **d**, image that is the convolution of a four-fold symmetric, noise-free model function, similar to those used previously<sup>13</sup> on incommensurate spin fluctuations in (214) and the response function of the instrument. The energy probed increases on moving diagonally (from bottom left) across the images, with the incommensurate peaks in **e** appearing at 22 and 26 meV. The counting times required to collect the images were 12, 16, 114, 63 h for images **b, c, e, f**, respectively, with the proton current of 170  $\mu\text{A}$ . The scattering was placed on an absolute scale using the elastic incoherent scattering from a vanadium standard under the same experimental conditions (see refs 16, 17, for example).



**Figure 2** Data obtained from cuts through the two-dimensional images in Fig. 1. **a**, Data obtained along  $(0,1)$  through the low-temperature resonance maximum at  $(1/2, 1/2)$  and 34 meV. **b–d**, Data obtained from cuts through the low-temperature incommensurate pattern. The coloured lines in Fig. 1a show the direction of the cuts. **b–d**, Data from cuts taken **b**, vertically through the corners of the square, and **c**, along the square edge; **d**, background. Solid lines are derived similarly from the model image in Fig. 1d. Horizontal bars represent the full-width at half-maximum of the experimental resolution function.

**Table 1 Comparison of parameters derived from neutron scattering data**

| Compound                                       | Ref.   | $x$  | $T$ (K) | Energy (meV) | $\delta$ (r.l.u.) | $\kappa_w$ ( $\text{\AA}^{-1}$ ) | $\chi''(\omega)$ ( $\mu_B^2 \text{eV}^{-1} \text{Cu}^{-1}$ ) |
|------------------------------------------------|--------|------|---------|--------------|-------------------|----------------------------------|--------------------------------------------------------------|
| $\text{La}_2\text{CuO}_4$                      | 16     | 0    | 295     | 20           | 0                 |                                  | $1.5 \pm 0.5$                                                |
| $\text{YBa}_2\text{Cu}_3\text{O}_{6.2}$        | 17     | 0    | 295     | 24           | 0                 |                                  | $1.8 \pm 0.4$                                                |
| $\text{La}_{1.9}\text{Sr}_{0.1}\text{CuO}_4$   | 14     | 0.1  | 8       | 8            | $0.21 \pm 0.02$   | $0.038 \pm 0.005$                |                                                              |
| $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$        |        | 0.1  | 13      | 24           | $0.21 \pm 0.02$   | $0.041 \pm 0.005$                | $2.5 \pm 0.5$                                                |
| $\text{La}_{1.86}\text{Sr}_{0.14}\text{CuO}_4$ | 13, 16 | 0.14 | 35      | 15           | $0.24 \pm 0.004$  | $0.052 \pm 0.002$                | $4 \pm 1$                                                    |
| $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$        |        | 0.1  | 13      | 34           | 0                 |                                  | $9.5 \pm 2$                                                  |

$x$  is the hole doping expressed in the units of electrons per planar Cu atom,  $\delta$  the incommensurability parameter in reciprocal lattice units,  $\kappa_w$  the half-width at half-maximum of the incommensurate peaks which reflects the inverse correlation length (related to but not equal to  $\kappa(\omega)$  in ref. 13), and  $\chi''(\omega)$  the wavevector-integrated susceptibility. Reference 14 defines  $\delta$  as a factor of 2 smaller. The wavevector-integrated susceptibility is expressed per planar Cu.

centred at  $(1/2, 1/2)$  which, although it is already present at  $T_c$ , grows substantially on cooling into the superconducting state.

Figure 1e and f shows the dramatically different patterns obtained at  $24.5 \pm 2.5$  meV: frame e reveals that the scattering at 13 K essentially disappears at  $(1/2, 1/2)$  and acquires maxima at the four positions displaced by  $(0, \pm\delta/2)$  and  $(\pm\delta/2, 0)$  from  $(1/2, 1/2)$ —the orientation of the pattern is clearly the same as for (214); frame f shows a measurement at 65 K that demonstrates that the pattern is broader at the higher temperature. In Fig. 2 we plot cuts through the maps along the lines shown in Fig. 1a. What emerges again is that the 24-meV low-temperature scattering can be visualized as a square box with subsidiary maxima at the corners and relatively little scattering at the commensurate position  $(1/2, 1/2)$ . The intensities are given as absolute scattering cross-sections and the solid lines are from a global fit to the same model used previously to analyse (214) neutron-scattering results<sup>13</sup>. A slight modification of the formalism was included which allows for an anisotropic peak shape and results in a better fit. Figure 1d shows how the fitting function appears in the two-dimensional reciprocal space for the  $\text{CuO}_2$  planes.

Scattering identified as stemming from the lattice vibrations (phonons) with energies near 20 meV is visible in the lower left part of the pattern in Fig. 1f. Some phonon scattering is also apparent at low temperatures, but it does not alter the shape of the pattern from the magnetic scattering. The present experiment is unique not only for its use of two-dimensional position-sensitive detection, but also because of its small phonon background. In particular, the evacuated and obstruction-free flight paths of the instrument permitted data collection at low scattering angles, resulting in low neutron momentum transfer,  $Q$ . Minimizing  $Q$  is the most efficient way to minimize phonon and multiphonon scattering, whose power varies as  $Q^2$  and  $Q^4$ , respectively. Although earlier experiments have looked near wavevectors such as  $(1/2, 3/2)$ —equivalent to  $(1/2, 1/2)$ —the actual net neutron momentum transfer has generally been  $\sim 2.73 \text{ \AA}^{-1}$ , whereas here  $Q \approx 1.47 \text{ \AA}^{-1}$ .

Table 1 lists parameters not only for our (123) sample at the two energy values probed, but also for (214) crystals examined previously. Parameters determined from the scattering patterns are the peak positions  $\delta$ , the inverse coherence length  $\kappa_w$ , given by the width of the features in the pattern, and the wavevector-integrated spectral weight  $\chi''(\omega)$ . Several conclusions emerge from Table 1, the most important being that the wavevectors for the incommensurate fluctuations in Fig. 1 are indistinguishable from those for similarly doped<sup>14</sup> single-layer (214) materials<sup>1,15</sup>. In addition, the coherence lengths of the incommensurate excitations are similar for similar doping levels and energy transfers scaled by  $T_c$  in the (214) and (123) materials. Furthermore, the integrated intensities of the incommensurate peaks, when measured at the same scaled energies, are also comparable<sup>16</sup>. They are also not far from the integrated spin wave intensities, measured at the same energies for the undoped antiferromagnetic parent compounds<sup>17</sup>. In contrast, the resonance in (123) is substantially more intense than the spin waves. A final point of resemblance between the spin fluctuations in (123) and (214) is that cooling into the superconducting phase appears to narrow the incommensurate peaks, and thus increases the magnetic

coherence. In our experiment, this effect is most clearly manifested by the poorer definition of the low-intensity 'hole' around  $(1/2, 1/2)$  at 65 K (Fig. 1f) compared with at 13 K (Fig. 1e); in previous work on slightly differently doped (214), it made itself apparent in better defined, enhanced incommensurate peaks<sup>18</sup>.

We have discovered that for the same hole doping, the low-frequency, subresonance, magnetic fluctuations in  $(123)\text{O}_{6.6}$  are remarkably similar to those in superconducting (214). Not only are their wavevectors and coherence lengths essentially the same, but their spectral weights are comparable. Furthermore, the fluctuations in both compounds share the phenomenon of increased coherence in the superconducting phase. The consequences for our understanding of the copper oxide superconductors are clear—the magnetic fluctuations at low frequencies are beginning to achieve the same universality across widely different materials with different values of  $T_c$ , which characterizes many of their macroscopic properties. As for the bulk characteristics, the controlling factor seems only to be the hole density. Thus all electronic structure, and other detailed effects not connected with  $\text{CuO}_2$  planes with a particular carrier density, appears to drop out at low frequencies, even for short-wavelength spin fluctuations. This represents a huge simplification of the high- $T_c$  problem and was not obvious from previous magnetic resonance, magnetic susceptibility, and neutron data. □

Received 29 April; accepted 2 July 1998.

- Cheong, S.-W. *et al.* Incommensurate magnetic fluctuations in  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ . *Phys. Rev. Lett.* **67**, 1791–1794 (1991).
- Mason, T. E., Aeppli, G. & Mook, H. A. Magnetic dynamics of superconducting  $\text{La}_{1.86}\text{Sr}_{0.14}\text{CuO}_4$ . *Phys. Rev. Lett.* **68**, 1414–1417 (1992).
- Thurston, T. R. *et al.* Low-energy incommensurate spin excitations in superconducting  $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$ . *Phys. Rev. B* **46**, 9128–9131 (1992).
- Rossat-Mignod, J. *et al.* Neutron scattering study of the  $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$  system. *Physica* **185** C, 86–92 (1991).
- Tranquada, J. M., Gehring, P. M., Shirane, G., Sato, M. & Sato, M. Neutron-scattering study of the dynamical spin susceptibility in  $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ . *Phys. Rev. B* **46**, 5561–5575 (1992).
- Sternlieb, B. J., Tranquada, J. M., Shirane, G., Sato, M. & Sato, M.  $q$  dependence of the dynamic susceptibility  $\chi''(q, \omega)$  in superconducting  $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$  ( $T_c = 46$  K). *Phys. Rev. B* **50**, 12915–12919 (1994).
- Vaknin, D. *et al.* Antiferromagnetism in  $\text{La}_2\text{CuO}_{4-x}$ . *Phys. Rev. Lett.* **58**, 2802–2805 (1987).
- Tranquada, J. M. *et al.* Neutron-diffraction determination of antiferromagnetic structure of Cu ions in  $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$  with  $x = 0.0$  and  $0.15$ . *Phys. Rev. Lett.* **60**, 156–159 (1988).
- Wells, B. O. *et al.* Incommensurate spin fluctuations in high-transition temperature superconductors. *Science* **277**, 1067–1071 (1997).
- Dai, P., Yethiraj, M., Mook, H. A., Lindemer, T. B. & Doğan, F. Magnetic dynamics in underdoped  $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ : direct observations of a superconducting gap. *Phys. Rev. Lett.* **77**, 5425–5428 (1996).
- Fong, H. F., Keimer, B., Milius, D. L. & Aksay, I. A. Superconducting-induced anomalies in the spin excitation spectra of underdoped  $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ . *Phys. Rev. Lett.* **78**, 713–716 (1997).
- Dai, P., Mook, H. A. & Doğan, F. Incommensurate magnetic fluctuations in  $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$ . *Phys. Rev. Lett.* **80**, 1738–1741 (1998).
- Aeppli, G., Mason, T. E., Hayden, S. M., Mook, H. A. & Kulda, J. Nearly singular magnetic fluctuations in the normal state of a high- $T_c$  cuprate superconductor. *Science* **278**, 1432–1435 (1997).
- Yamada, K. *et al.* Doping dependence of the spatially modulated dynamical spin correlations and the superconducting-transition temperature in  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ . *Phys. Rev. B* **57**, 6165–6172 (1998).
- Tallon, J. L., Bernhard, C., Shaked, H., Hitterman, R. L. & Jorgensen, J. D. Generic superconducting phase behavior in high- $T_c$  cuprates:  $T_c$  variation with hole concentration in  $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ . *Phys. Rev. B* **51**, 12911–12914 (1995).
- Hayden, S. M. *et al.* Comparison of the high-frequency magnetic fluctuations in insulating and superconducting  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ . *Phys. Rev. Lett.* **76**, 1344–1347 (1996).
- Hayden, S. M., Aeppli, G., Perring, T. G., Mook, H. A. & Doğan, F. High-frequency spin waves in  $\text{YBa}_2\text{Cu}_3\text{O}_{6.15}$ . *Phys. Rev. B* **54**, R9605–R9608 (1996).
- Mason, T. E., Schröder, A., Aeppli, G., Mook, H. A. & Hayden, S. M. New magnetic coherence effect in superconducting  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ . *Phys. Rev. Lett.* **77**, 1604–1607 (1996).

**Acknowledgements.** We thank R. Coldea, D. A. Tennant, C. Frost and T. B. Lindemer for technical assistance. Support at Oak Ridge was received from the US DOE.

Correspondence and requests for materials should be addressed to H.A.M. (e-mail: ham@ornl.gov).



# Solid-state dye-sensitized mesoporous TiO<sub>2</sub> solar cells with high photon-to-electron conversion efficiencies

U. Bach\*, D. Lupo†‡, P. Comte\*, J. E. Moser\*, F. Weissörtel§, J. Salbeck§, H. Spreitzer† & M. Grätzel\*

\* Institute of Photonics and Interfaces, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland

† Hoechst Research & Technology Deutschland GmbH & Co. KG, Industriepark Höchst, D-65926 Frankfurt, Germany

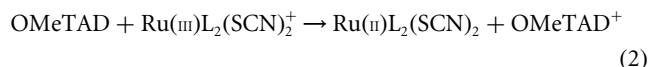
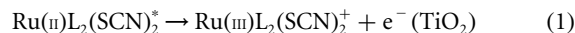
§ Max-Planck-Institut für Polymerforschung, D-55128 Mainz, Germany

Solar cells based on dye-sensitized mesoporous films of TiO<sub>2</sub> are low-cost alternatives to conventional solid-state devices<sup>1</sup>. Impressive solar-to-electrical energy conversion efficiencies have been achieved with such films when used in conjunction with liquid electrolytes<sup>2</sup>. Practical advantages may be gained by the replacement of the liquid electrolyte with a solid charge-transport material. Inorganic p-type semiconductors<sup>3,4</sup> and organic materials<sup>5–9</sup> have been tested in this regard, but in all cases the incident monochromatic photon-to-electron conversion efficiency remained low. Here we describe a dye-sensitized heterojunction of TiO<sub>2</sub> with the amorphous organic hole-transport material 2,2',7,7'-tetrakis(N,N-di-p-methoxyphenyl-amine)9,9'-spirobifluorene (OMeTAD; refs. 10 and 11). Photoinduced charge-carrier generation at the heterojunction is very efficient. A solar cell based on OMeTAD converts photons to electric current with a high yield of 33%.

The hole conductor contains a spiro-centre (a tetrahedral carbon linking two aromatic moieties) which is introduced in order to improve the glass-forming properties and prevent crystallization of

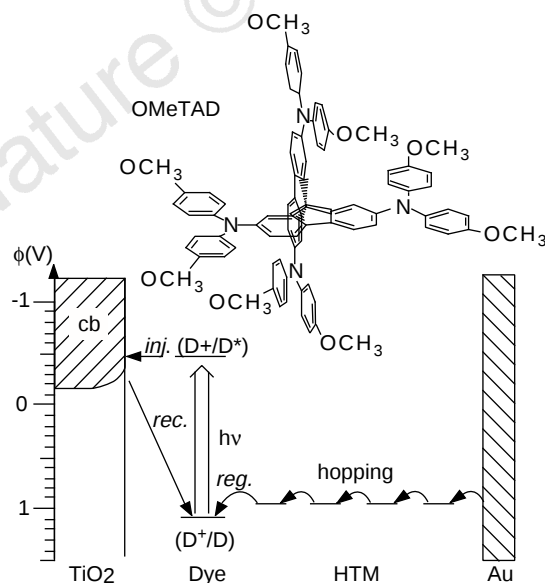
the organic material. Its glass transition temperature of  $T_g = 120^\circ\text{C}$ , measured by differential scanning calorimetry, is much higher than that of the widely used hole conductor TPD (N,N'-diphenyl-N,N'-bis(3-methylphenyl)4,4'-diamine;  $T_g = 62^\circ\text{C}$ ). Crystallization is undesirable as it would impair the formation of a good contact between the mesoporous surface of the TiO<sub>2</sub> and the hole conductor. The methoxy groups are introduced in order to match the oxidation potential of the hole-transport medium (HTM) to that of the sensitizer Ru(II)L<sub>2</sub>(SCN)<sub>2</sub> (where L is 4,4'-dicarboxy-2,2'-bipyridyl), used in this study.

Figure 1 shows a scheme for the electron-transfer processes occurring at the dye-sensitized heterojunction. Visible-light absorption by the sensitizer is followed by electron transfer to the conduction band of TiO<sub>2</sub>. The dye is regenerated by hole injection into the HTM. The TiO<sub>2</sub> conduction-band electrons, as well as the holes in the HTM, are subsequently transported by electronic conduction to the contact electrodes. Pulsed nanosecond laser photolysis was used in conjunction with time-resolved absorption spectroscopy to scrutinize the dynamics of the photoinduced charge separation process. Figure 2 shows the transient absorption spectrum of a dye-sensitized mesoporous TiO<sub>2</sub> film in the absence and presence of OMeTAD, measured 50 ns after laser excitation. In the absence of OMeTAD, dye bleaching at ~500 nm is observed and a broad positive transient absorption appears above 600 nm due to the absorption of the oxidized dye Ru(III)L<sub>2</sub>(SCN)<sub>2</sub><sup>+</sup> and of the TiO<sub>2</sub> conduction-band electrons. Electron injection proceeds in the femtosecond domain<sup>12</sup>, while the subsequent recapture of injected electrons by the oxidized dye takes several microseconds. In the presence of OMeTAD, the bleaching signal disappears. Instead, the transient absorption rises vertically within the laser pulse. Comparison of the transient spectra obtained in the presence of OMeTAD with the absorption band of chemically oxidized OMeTAD confirms that the species giving rise to the new spectral feature is the radical cation OMeTAD<sup>•+</sup>. Apparently, electron injection from the excited sensitizer into TiO<sub>2</sub> is immediately followed by regeneration of the dye via hole transfer to OMeTAD, as shown below:

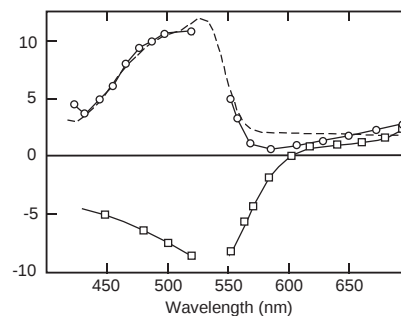


The process shown in equation (2) was too fast to be monitored with the laser equipment employed, setting an upper limit of 40 ns for the hole-transfer time.

A blank experiment was performed using mesoporous Al<sub>2</sub>O<sub>3</sub> films instead of TiO<sub>2</sub> as a support for the Ru(II)L<sub>2</sub>(SCN)<sub>2</sub> sensitizer; the results showed that hole transfer from the excited state of the dye



**Figure 1** Scheme for the electron-transfer processes (inj., injection; reg., regeneration; rec., recapture; hopping) occurring in the dye-sensitized heterojunction. Also shown are the approximate redox potentials and band energies of the different components.



**Figure 2** Absorption spectra from time-resolved laser photolysis experiments. Shown are the transient absorption spectra of a dye-sensitized mesoporous TiO<sub>2</sub> film in the absence (squares, contact medium propylene carbonate) and in the presence (circles) of solid OMeTAD, 50 ns after excitation at 532 nm. For comparison, the absorption spectrum of chemically oxidized OMeTAD in chlorobenzene:acetonitrile = 90:10 is also shown (dashed line, arbitrary units).

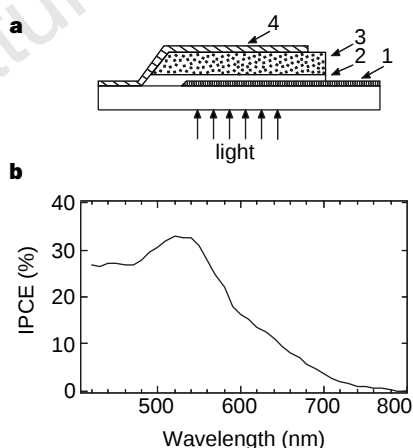
‡ Present address: Materials Science Laboratories, Sony International (Europe) GmbH, Stuttgarter Strasse 106, 70736 Fellbach, Germany.

to the OMeTAD does not contribute significantly to the photo-induced charge-separation phenomena observed.

The photovoltaic performance of the dye-sensitized heterojunction was studied by means of sandwich-type cells, shown schematically in Fig. 3a. The working electrode consisted of conducting glass (F-doped  $\text{SnO}_2$ , sheet resistance  $10\ \Omega$  per square) onto which a compact  $\text{TiO}_2$  layer was deposited by spray pyrolysis<sup>13</sup>. This avoids direct contact between the HTM layer and the  $\text{SnO}_2$  which would short-circuit the cell. A  $4.2\text{-}\mu\text{m}$ -thick mesoporous film of  $\text{TiO}_2$  was deposited by screen printing onto the compact layer<sup>14</sup>, and derivatized with  $\text{Ru}(\text{II})\text{L}_2(\text{SCN})_2$  by adsorption from acetonitrile. The HTM was introduced into the mesopores by spin-coating a solution of OMeTAD in chlorobenzene onto the  $\text{TiO}_2$  film, and subsequent evaporation of the solvent. A semi-transparent gold back contact was evaporated on top of the hole conductor under vacuum.

Figure 3b shows the photocurrent action spectrum of a typical cell under short-circuit conditions. The given values are not corrected for reflection and absorption losses of the conducting glass, which are estimated to be at least 15% in the visible region of the spectrum. The spectrum closely matches the absorption spectrum of the dye, confirming that the observed photocurrent arises from electron injection by the sensitizer. The maximum value of the incident photo-to-electron conversion efficiency (IPCE) is 33%, which is more than two orders of magnitude larger than the previously reported value for a similar dye-sensitized solid heterojunction<sup>9</sup> and only a factor of  $\sim 2$  lower than with liquid electrolytes<sup>2</sup>.

The coating solution used for the device in Fig. 3b contained  $0.33\text{ mM}$   $\text{N}(\text{PhBr})_3\text{SbCl}_6$  and  $15\text{ mM}$   $\text{Li}[(\text{CF}_3\text{SO}_2)_2\text{N}]$  in addition to  $0.17\text{ M}$  OMeTAD. In the absence of these additives, the maximum IPCE was only 5%.  $\text{N}(\text{PhBr})_3\text{SbCl}_6$  acts as a dopant, introducing free charge carriers in the HTM by oxidation, as confirmed by spectro-electrochemical measurements. Partial oxidation of OMeTAD by  $\text{N}(\text{PhBr})_3\text{SbCl}_6$  is a convenient way to control the dopant level<sup>15</sup>. On adding  $\text{N}(\text{PhBr})_3\text{SbCl}_6$  to a solution of OMeTAD in chlorobenzene, the radical cation  $\text{OMeTAD}^+$  is instantly formed. The spectral features of  $\text{OMeTAD}^+$  remained unchanged during solvent evaporation and glass formation, except for a small hypochromic shift. No subsequent absorption changes were detectable over several weeks, confirming the temporal stability of  $\text{OMeTAD}^+$  in the HTM.

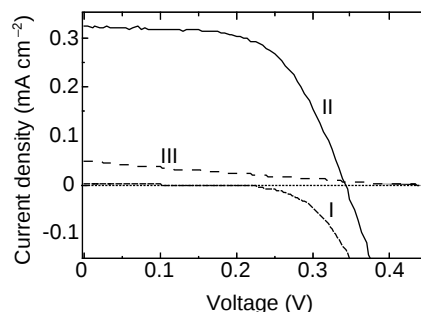


**Figure 3** Structure and spectral response of the photovoltaic devices. **a**, Structure 1, conducting F-doped  $\text{SnO}_2$ -coated glass; 2, compact  $\text{TiO}_2$  layer; 3, dye-sensitized heterojunction; 4, gold electrode. **b**, Photocurrent action spectrum for a dye-sensitized heterojunction, the structure of which is shown above. The IPCE value corresponds to the number of electrons generated by monochromatic light in the external circuit, divided by the number of incident photons. The  $4.2\text{-}\mu\text{m}$ -thick mesoporous  $\text{TiO}_2$  film was sensitized with  $\text{Ru}(\text{II})\text{L}_2(\text{SCN})_2$ , spin-coated with a solution of  $0.17\text{ M}$  OMeTAD,  $0.33\text{ mM}$   $\text{N}(\text{PhBr})_3\text{SbCl}_6$  and  $15\text{ mM}$   $\text{Li}[(\text{CF}_3\text{SO}_2)_2\text{N}]$  in chlorobenzene with 5% acetonitrile added.

The second additive,  $\text{Li}[(\text{CF}_3\text{SO}_2)_2\text{N}]$ , is a source of  $\text{Li}^+$  ions, which are known to be potential-determining for  $\text{TiO}_2$  (ref. 16). Along with the protons from the carboxylic acid groups of  $\text{Ru}(\text{II})\text{L}_2(\text{SCN})_2$ , they confer a positive charge on the surface of the oxide. As the sensitizer is negatively charged a local electrostatic field is produced, assisting electron injection into the  $\text{TiO}_2$  while retarding recapture of the electron by the oxidized dye. The lithium salt may also compensate for space-charge effects. Under light illumination of the heterojunction, a net positive space charge is expected to be formed in the HTM, inducing a local field that impairs current flow. The lithium salt could screen this field, thereby eliminating the space-charge control of the photocurrent. Improvement of the photovoltaic performance of dye-sensitized heterojunctions by immersion in  $\text{LiClO}_4$  solutions was also reported by Murakoshi *et al.*<sup>5</sup>.

Figure 4 shows current-density/voltage curves employing the device structure shown in Fig. 3a. Curves I and II were obtained with hole conductor containing both the  $\text{N}(\text{PhBr})_3\text{SbCl}_6$  dopant and the  $\text{Li}[(\text{CF}_3\text{SO}_2)_2\text{N}]$  salt, whereas these additives were absent for curve III. Curve I was measured in the dark, whereas II and III were obtained under light illumination. The device that contains the hole conductor without additives performs poorly, the conversion yield being only 0.04% at a white-light illumination of  $9.4\text{ mW cm}^{-2}$ . Addition of the dopant and  $\text{Li}^+$  salt increases the overall conversion efficiency to 0.74%. Under full sunlight ( $100\text{ mW cm}^{-2}$ , air mass 1.5), the short-circuit photocurrent density reached  $3.18\text{ mA cm}^{-2}$ , a value which is unprecedented for solar cells based on organic solids. Further improvement of the photovoltaic performance is expected, as many parameters of the cell assembly have not yet been optimized. Preliminary stability tests performed over 80 h using the visible output of a 400 W Xe lamp showed that the photocurrent was stable within  $\pm 20\%$ , while the open-circuit voltage and the fill factor (see Methods) increased. The total charge passed through the cell during illumination was  $300\text{ C cm}^{-2}$ ; corresponding to turnover numbers of about 8,400 and 60,000 for the OMeTAD and the dye, respectively. This shows that the hole conductor can sustain photovoltaic operation without significance degradation.

From the present findings, the concept of dye-sensitized heterojunctions emerges as a very interesting and viable option for future



**Figure 4** Current-density/voltage characteristics. Shown are characteristics of the same device as in Fig. 3, obtained in the dark (I) and under white-light illumination at  $9.4\text{ mW cm}^{-2}$  (II). The spectral distribution corresponded to global air mass 1.5 corrected for spectral mismatch. The short-circuit current density was  $0.32\text{ mA cm}^{-2}$ , the open-circuit voltage 342 mV and the fill factor 62% corresponding to an overall conversion efficiency of 0.74%. For comparison, the photocurrent-density/voltage characteristic of a cell containing no  $\text{N}(\text{PhBr})_3\text{SbCl}_6$  or  $\text{Li}[(\text{CF}_3\text{SO}_2)_2\text{N}]$  is also shown (III).

low-cost solid-state solar cells. Photodiodes based on interpenetrating polymer networks of poly(phenylenevinylene) derivatives<sup>17,18</sup> present a related approach. The main difference to our system is that at least one component of the polymer network needs to function simultaneously as an efficient light absorber and a good charge-transport material. The dye-sensitized heterojunction cell offers a greater flexibility, as the light absorber and charge-transport material can be selected independently to obtain optimum solar-energy harvesting and high photovoltaic output. □

## Methods

**Compounds.** OMeTAD was pure according to <sup>1</sup>H-NMR and HPLC analysis. The synthesis will be reported elsewhere. Ru(II)L<sub>2</sub>(SCN)<sub>2</sub> was prepared as previously described<sup>2</sup>.

**Transient absorption spectroscopy.** This was carried out with a Nd-YAG laser as excitation light source, producing a 6-ns pulse at 532 nm of typically 1 mJ cm<sup>-2</sup>, with a repetition frequency of 30 Hz. The probe light was provided by a Xe lamp, which was spectrally narrowed by cut-off and interference filters before passing the device. A monochromator combined with a photomultiplier was used as detection system. A Tektronix 524 TDS oscilloscope was used to record and store the data. For the laser experiments, dye-sensitized mesoporous semiconductor films were deposited on ordinary glass.

**Photocurrent-voltage characteristics.** These were measured with a Keithley 2400 Source Meter and a 400 W Xe lamp. A Schott KG3 filter was used in order to approach the spectral distribution of the lamp to air mass 1.5 G. The light intensity was regulated to the desired energy output by using a silicon solar cell, calibrated at the ISE-Fraunhofer Institut in Freiburg Germany. Efficiencies were corrected for the spectral mismatch. The fill factor (FF) is defined as  $FF = V_{opt} I_{opt} / I_{sc} V_{oc}$ , where  $V_{opt}$  and  $I_{opt}$  are respectively current and voltage for maximum power output, and  $I_{sc}$  and  $V_{oc}$  are the short-circuit current and open-circuit voltage, respectively.

Received 8 May; accepted 13 July 1998.

- O'Regan, B. & Grätzel, M. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO<sub>2</sub> films. *Nature* **353**, 737–739 (1991).
- Nazeeruddin, M. K. *et al.* Conversion of light to electricity by cis-X<sub>2</sub>bis(2,2'-bipyridyl)-4,4'-dicarboxylate) ruthenium(II) charge-transfer sensitizers (X = Cl<sup>-</sup>, Br<sup>-</sup>, I<sup>-</sup>, CN<sup>-</sup> and SCN<sup>-</sup>) on nanocrystalline TiO<sub>2</sub> electrodes. *J. Am. Chem. Soc.* **115**, 6382–6390 (1993).
- O'Regan, B. & Schwarz, D. T. Large enhancement in photocurrent efficiency caused by UV illumination of the dye-sensitized heterojunction TiO<sub>2</sub>/RuL<sub>2</sub>NCS/CuSCN: initiation and potential mechanism. *Chem. Mater.* **10**, 1501–1509 (1998).
- Tennakone, K., Kumara, G. R. A., Kumarasinghe, A. R., Wijayantha, K. G. U. & Sirimanne, P. M. Dye-sensitized nano-porous solid-state photovoltaic cell. *Semicond. Sci. Technol.* **10**, 1689–1693 (1995).
- Murakoshi, K., Kogure, R. & Yanagida, S. Solid state dye-sensitized TiO<sub>2</sub> solar cell with polypyrrole as hole transport layer. *Chem. Lett.* **5**, 471–472 (1997).
- Bach, U. *et al.* Ultrafast hole injection from dye molecules into an organic hole conductor for dye sensitized solid state solar cells. *Abstract Book, Bayreuth Polymer & Materials Research Symposium*, P28 (Bayreuth, 1997).
- Weissörtel, F. *Amorphous niedermolekulare Ladungstransportmaterialien für nanokristalline Solarzellen*. Thesis, Univ. Regensburg (1996).
- Grätzel, M. in *Future Generation Photovoltaic Technologies* Vol. 404 (ed. McConnell, R.) 119–126 (Am. Inst. Phys., Denver, 1997).
- Hagen, J. *et al.* Novel hybrid solar cells consisting of inorganic nanoparticles and an organic hole transport material. *Synth. Met.* **89**, 215–220 (1997).
- Salbeck, J., Weissörtel, F. & Bauer, J. Spiro linked compounds for use as active materials in organic light emitting diodes. *Macromol. Symp.* **125**, 121–132 (1997).
- Salbeck, J., Yu, N., Bauer, J., Weissörtel, F. & Bestgen, H. Low molecular organic glasses for blue electroluminescence. *Synth. Met.* **91**, 209–215 (1997).
- Tachibana, Y., Moser, J. E., Grätzel, M., Klug, D. R. & Durrant, J. R. Subpicosecond interfacial charge separation in dye-sensitized nanocrystalline titanium dioxide films. *J. Phys. Chem.* **100**, 20056–20062 (1996).
- Kavan, L. & Grätzel, M. Highly efficient semiconducting TiO<sub>2</sub> photoelectrodes prepared by aerosol pyrolysis. *Electrochim. Acta* **40**, 643–652 (1995).
- Barbé, C. J. *et al.* Nanocrystalline titanium oxide electrodes for photovoltaic applications. *J. Am. Ceram. Soc.* **80**, 3157–3171 (1997).
- Abkowitz, M. & Pai, D. M. Comparison of the drift mobility measured under transient and steady-state conditions in a prototypical hopping system. *Phil. Mag.* **B 53**, 192–216 (1986).
- Enright, B., Redmond, G. & Fitzmaurice, D. Spectroscopic determination of flat-band potentials for polycrystalline TiO<sub>2</sub> electrodes in mixed-solvent systems. *J. Phys. Chem.* **97**, 1426–1430 (1994).
- Halls, J. J. M. *et al.* Efficient photodiodes from interpenetrating polymer networks. *Nature* **376**, 498–500 (1995).
- Yu, G., Gao, J., Hummelen, J. C., Wudl, F. & Heeger, A. J. Polymer photovoltaic cells: enhanced efficiencies via a network of internal donor acceptor heterojunctions. *Science* **270**, 1789–1791 (1995).

**Acknowledgement.** This work was supported by the Swiss National Science Foundation and the European Joule III programme (OFES).

Correspondence and requests for materials should be addressed to M.G. (e-mail: michael.gratzel@epfl.ch).

# Accumulation of persistent organochlorine compounds in mountains of western Canada

Jules M. Blais<sup>\*†</sup>, David W. Schindler<sup>\*</sup>, Derek C. G. Muir<sup>†‡</sup>, Lynda E. Kimpe<sup>§</sup>, David B. Donald<sup>||</sup> & Bruno Rosenberg<sup>¶</sup>

<sup>\*</sup> Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2E9

<sup>‡</sup> Department of Fisheries and Oceans, Freshwater Institute, 501 University Crescent, Winnipeg, Manitoba, Canada R3T 2N6

<sup>§</sup> Public Health Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G3

<sup>||</sup> Environment Canada, Room 300 Park Plaza, 2365 Albert Street, Regina, Saskatchewan, Canada S4P 4K1

<sup>¶</sup> Freshwater Institute, Winnipeg, Manitoba, Canada R3T 2N6

Persistent, semi-volatile organochlorine compounds, including toxic industrial pollutants and agricultural pesticides, are found everywhere on Earth, including in pristine polar and near-polar locations<sup>1–4</sup>. Higher than expected occurrences of these compounds in remote regions are the result of long-range transport in the atmosphere, precipitation and 'cold condensation'—the progressive volatilization in relatively warm locations and subsequent condensation in cooler environments<sup>3,4</sup> which leads to enhanced concentrations at high latitudes. The upper reaches of high mountains are similar to high-latitude regions in that they too are characterized by relatively low average temperatures, but the accumulation of organochlorine compounds as a function of altitude has not yet been documented. Here we report organochlorine deposition in snow from mountain ranges in western Canada that show a 10- to 100-fold increase between 770 and 3,100 m altitude. In the case of less-volatile compounds, the observed increase by a factor of 10 is simply due to a 10-fold increase in snowfall over the altitude range of the sampling sites. In the case of the more-volatile organochlorines, cold-condensation effects further enhance the concentration of these compounds with increasing altitude. These findings demonstrate that temperate-zone mountain regions, which tend to receive

**Table 1 Correlation between organochlorine concentrations in snow and site elevations**

| Compound          | Correlation coefficient | Vapour pressure (Pa)                                   |
|-------------------|-------------------------|--------------------------------------------------------|
| α-HCH             | 0.85*                   | 0.1                                                    |
| Heptachlorepoxyde | 0.75*                   | 0.1                                                    |
| γ-HCH             | 0.73*                   | 0.03                                                   |
| Dieldrin          | 0.42*                   | 0.016                                                  |
| Endosulphan-I     | 0.76*                   | 0.008                                                  |
| c-Chlordane       | 0.42*                   | 0.003                                                  |
| t-Chlordane       | 0.34                    | 0.003                                                  |
| p,p'-DDT          | 0.00                    | 0.0001                                                 |
| PCBs              |                         |                                                        |
| Σ Dichloro-       | 0.54*                   | 0.2 (0.008–0.60)                                       |
| Σ Trichloro-      | 0.53*                   | 0.04 (0.003–0.022)                                     |
| Σ Tetrachloro-    | 0.00                    | 0.006 (0.003–0.10)                                     |
| Σ Pentachloro-    | 0.00                    | 0.001 (0.0003–0.009)                                   |
| Σ Hexachloro-     | 0.11                    | 0.0002 (7 × 10 <sup>-4</sup> – 0.012)                  |
| Σ Heptachloro-    | 0.17                    | 3 × 10 <sup>-4</sup> (2.7 × 10 <sup>-5</sup> – 0.0015) |

Correlation coefficients (r) are shown for organochlorine concentrations (ng l<sup>-1</sup>) in snow and site elevation (m.a.s.l.) for the equation  $\text{conc.} = ae^{b \text{Elev}}$ , where a and b are fitted constants. Asterisks show significance at P ≤ 0.05, for 19 degrees of freedom. Sub-cooled liquid vapour pressures are included for pesticides at 20°C (ref. 21) and PCBs at 25°C (ref. 22). Published vapour pressures vary considerably, so these values represent mean reported values for all PCBs in that class. Ranges of published vapour pressures for each PCB category are shown in brackets. Only compounds with mean sample concentrations that were ten times higher than blanks were considered.

<sup>†</sup> Present addresses: Department of Biology, University of Ottawa, 30 Marie Curie Street, PO Box 450 Stn. A, Ottawa, Ontario, Canada K1N 6N5 (J.M.B.); Environment Canada, 867 Lakeshore Road, Burlington, Ontario, Canada L7R 4A6 (D.C.G.M.).



high levels of precipitation while being close to pollutant sources, are particularly susceptible to the accumulation of semivolatile organochlorine compounds.

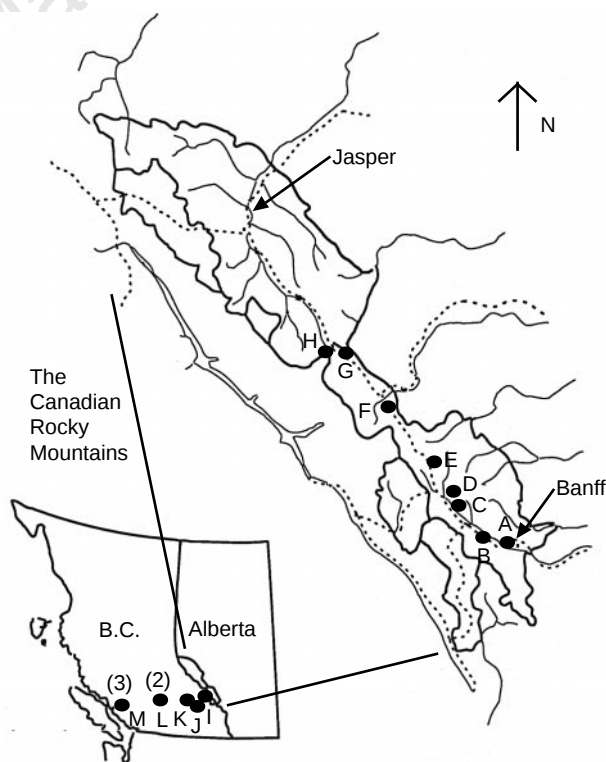
The present study was prompted by the observation that fish from high-altitude lakes in the Canadian Rocky Mountains contain high concentrations of chlorinated organic contaminants<sup>1,5,6</sup>. Point sources of pollutants to the lakes are not known and, in contrast to the conditions encountered in subarctic, low-altitude lakes<sup>7,8</sup>, the lakes in the Canadian Rockies are not characterized by unusually long food chains that would lead to high pollutant concentrations in fish as a result of extreme biomagnification<sup>6</sup>. In view of these considerations, atmospheric deposition appears to be the only process that can provide a plausible explanation for the observed contamination of fish in high-altitude Canadian lakes. Increased atmospheric deposition of organochlorines at high-elevation sites could be caused by the larger precipitation rates and reduced volatilization at high altitudes. (The tendency to volatilize decreases with temperature, and hence altitude.) In-cloud scavenging processes, which have been shown to increase the concentration of strong acids (H<sub>2</sub>SO<sub>4</sub> and HNO<sub>3</sub>) in precipitation reaching high-elevation sites in the eastern United States<sup>9</sup> and Switzerland<sup>10</sup>, may also contribute to the enhanced atmospheric deposition of organochlorines in the Canadian Rockies.

Snowpack samples (22) were collected in the vicinity of 15 Canadian meteorological stations in the coastal mountains of British Columbia as far inland as Banff, Alberta, and along the continental divide between Banff and Jasper, Alberta (Fig. 1). Snowpack samples were taken within a two-week period in late February and early March 1996, before any melting had occurred and when snowpacks were approaching their peak depth for the winter. Meteorological data from these sites indicated a decrease in mean monthly temperatures of 6 °C for every 1,000 m increase in elevation. Mean monthly temperatures in January for all sites

(except Snowdome, for which there are no data) ranged between -8 and -18 °C. All but one of these snowpacks represented the 1995/96 accumulation year only, because there is complete melting of snow in spring. The exception was the glacier Snowdome at 3,100 m above sea level (m.a.s.l.), which was sampled in March 1995. Although there is permanent snowpack at this site, annual layers can be easily identified. Our sample from that site represented the 1994/95 accumulation year. Snow was collected in specially designed sealed aluminium cans that were pre-rinsed with acetone, hexane and dichloromethane. Snowpack depth (in water equivalents) and snow density were determined at each site using snow cores at the time of sampling. Samples were slowly melted in the laboratory in airtight containers to suppress volatilization of analytes and 20 litres of water was pumped through a large-volume Goulden extractor where organochlorine residues were extracted into dichloromethane, along with internal standards to determine extraction efficiencies. Extracts were then purified into three fractions with 1.2% deactivated Florisil (8 g) and solvent-exchanged into iso-octane as described in other studies<sup>11,12</sup>. Samples were injected into a gas chromatograph equipped with a <sup>63</sup>Ni electron-capture detector. The main peaks were confirmed in selected samples by gas chromatography/mass spectrometry (HP 5971 MSD) using a 30-m DB-5 column with He carrier gas.

Contaminant concentrations increased with increasing altitude. Regression analysis showed that concentrations of heptachlor epoxide,  $\alpha$ -HCH (hexachlorocyclohexane),  $\gamma$ -HCH, *c*-chlordane, endosulphan-I and dieldrin, as well as the total concentration of di- and tri-chlorobiphenyls (PCBs), were positively correlated with site elevation (Table 1).

Organochlorine inventories in snow, defined as the concentration of the chemical in snow multiplied by snow depth in water equivalents, increased with altitude by factors of 10 to 100 over the elevation range studied (Fig. 2). The increase was due to a



**Figure 1** Sample sites for the analysis of organochlorines in snow. Numbers in brackets indicate multiple locations from that site. Dotted lines represent roads. Site names and elevations (m.a.s.l.) are as follows: A, Banff, 1,397; B, Sunshine, 2,186; C, Castle Mountain, 1,532, two reps; D, Lake Louise, 1,524; E, Bow summit,

2,011 three reps; F, Saskatchewan River Crossing, 1,402; G, Parker Ridge, 2,011; H, Snowdome glacier, 3,100; I, Donald, 770; J, Boulder Creek, 1,219; K, Wapta Lake, 1,890; L, Tod Mountain, two sites, 1,860 and 2,060, two reps each; M, Whistler, three sites, 1,625 with two reps, 1,660, and 1,850.

combination of the higher concentrations of organochlorines in snow at higher altitudes (Table 1, Fig. 2), and increasing precipitation with altitude. Snow depth (cm water equivalent) was strongly correlated with both site elevation and proximity to the Pacific coast. Multiple regression analysis showed that both elevation (in m.a.s.l.) and longitude ( $^{\circ}$ W) were significantly correlated ( $P < 0.05$ ) with snow depth (cm water equivalent) according to the following equation, with 82% of among-site variance explained ( $R^2$ ):

$$\text{Snow depth} = (0.037 \text{ elev.}) + (6.611 \text{ long.}) - 801 \quad (1)$$

The depth of snowpacks in water equivalents ranged between 12 and 116 cm over the study sites. The distribution of more volatile organochlorines in snowpacks (Fig. 2) showed a gradual increase with elevation at lower altitudes, followed by a sharp increase beyond 2,000 m.a.s.l.

As predicted by the cold-condensation hypothesis, the increase in snow concentrations of PCBs with altitude was greater for the more-volatile, less-chlorinated di- and tri-chlorobiphenyl congeners than it was for heavier, less-volatile forms (Fig. 3). A predominance of lower-molecular-weight PCBs would be expected to occur in colder regions owing to their higher liquid vapour pressures<sup>13</sup>. A similar relative enhancement of low-molecular-weight PCBs in the arctic relative to more southerly locations has been shown in lake sediments<sup>14</sup> and burbot liver<sup>15</sup>.

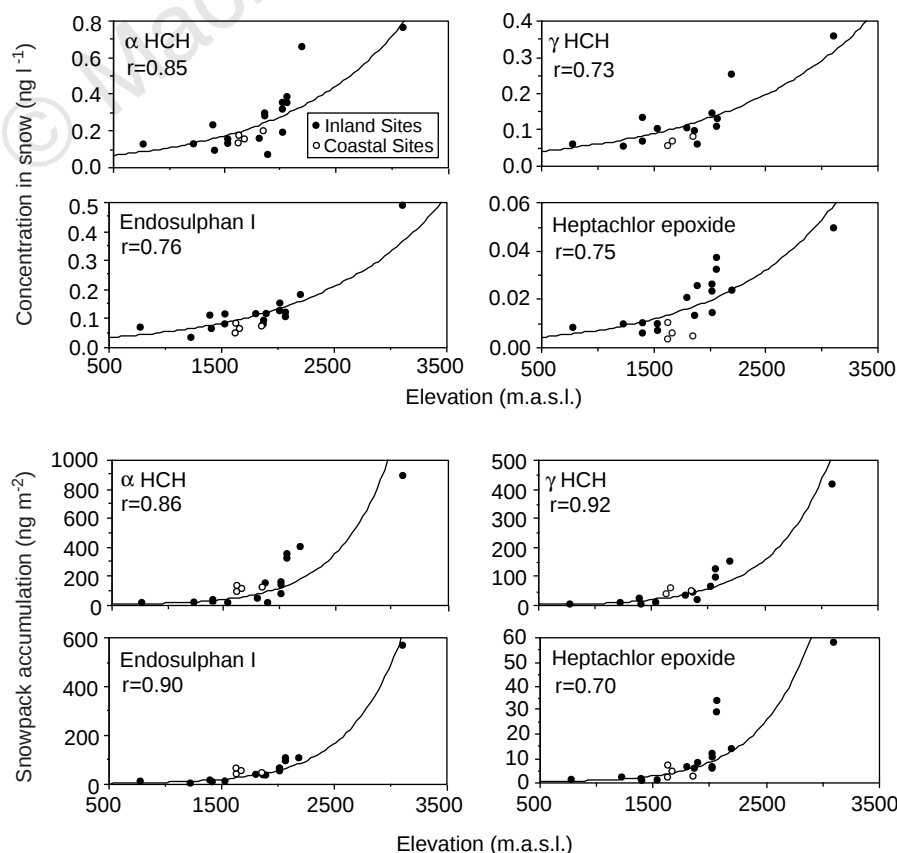
Considerable revolatilization of organochlorines from snowpacks takes place as a snowpack matures through the winter accumulation period. The greatest volatilization losses from snow on the Agassiz ice cap, Ellesmere island, were observed for  $\alpha$ - and  $\gamma$ -HCH, endosulphan-I and dieldrin<sup>16</sup>. These are the same compounds that we observed to become enriched in high-altitude snowpacks, indicating that losses due to revolatilization from snowpacks may be lower at higher elevation, consistent with the cold-condensation

hypothesis. Other investigators have found that more water-soluble constituents (for example, HCH) were drained with the first melt waters, whereas less-soluble compounds were released later with particulate matter<sup>17</sup>.

The organochlorines found in highest concentrations were the total concentration of HCH ( $\Sigma$ HCH)  $\Sigma$ PCBs and endosulphan-I. Among these organochlorines, only lindane ( $\gamma$ -HCH) and endosulphan are currently used in Canada and the USA. The ratio of  $\alpha$ -/ $\gamma$ -HCH isomers in the technical formulation of HCH ranges between 4 and 15 (ref. 18). In this study, the  $\alpha$ -/ $\gamma$  ratio ranges between 1.2 and 3.1, indicating that lindane is unlikely to be the only source of HCH to this region.

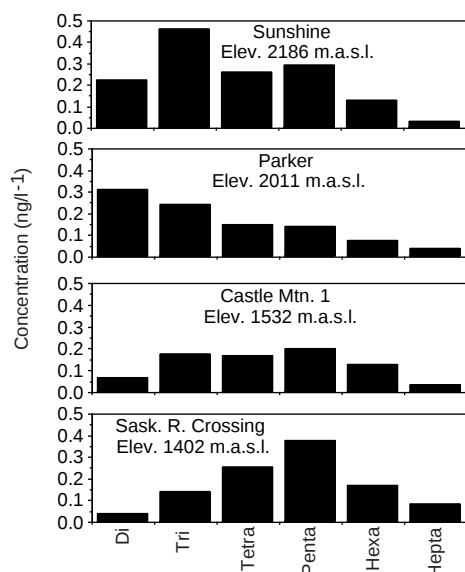
The inventories of the polychlorinated biphenyls ( $\Sigma$ PCBs) in the snowpacks tended to be lower in the coastal mountains of British Columbia than in the inland mountains of Alberta. Concentrations of PCBs in snow ranged between 0.30 and 2.17  $\text{ng l}^{-1}$ , which are similar to those measured previously in snowpacks from the high Arctic<sup>19</sup>, but PCB inventories in snow from our study region are much higher as a result of higher precipitation.

Our results show the potential for mountain regions to retain semivolatile organochlorine pesticides and PCBs. It is at present unclear how much of the contaminants deposited at high elevations may find their way into the surface waters at more populous low-altitude sites. Surface waters for many temperate and subtropical regions of western North America are derived largely from mountainous regions owing to high precipitation at high elevations. For example, the mountainous area of western Alberta, where our inland samples were taken, supplies 87% of the flow to the Saskatchewan River system, while constituting only 12% of the surface area of the basin<sup>20</sup>. Many peaks in the Rockies reach heights that exceed those from our study, and there is reason to believe that levels of organochlorines in snow would continue to increase at



**Figure 2** Organochlorine concentrations in snow and snowpack inventories for selected organochlorines as a function of elevation for all sites. All replicates are shown as separate points. Increased enrichment in concentration with elevation

was observed for more volatile compounds. Open circles, samples taken near the British Columbia coast (site M); filled circles, inland sites.



**Figure 3** Concentrations of PCBs, classed by number of chlorine atoms per molecule in snow from four of the study sites. Vapour pressures decrease with increasing chlorination (Table 1). We note the increase in the more-volatile, less-chlorinated PCBs at high elevation. Similar tendencies have been shown in lake sediments along latitudinal gradients<sup>14</sup>. Dichlorobiphenyls are PCB numbers 6,8/5; tri-17,18,19,16/32,24/27,22,25,26,28,31,33; tetra-41/47,40,42,44,45,46,47,48,49,52,56/60,64,70/76,74,82; penta-83,84/89,85,91,97,99,101,105,110,114,118; hexa-128,130/176,131,132,134,136,137,138,141,146,144/135,149,151,153,158; hepta-174,175,177,179,178/129,183,185,187.

higher elevations. For example, cities like Denver and Mexico City derive their water supply from snowmelt on mountains over 3,000 m high. They are also much closer to industrial and agricultural sources of contaminants in areas that reach summer temperatures 5–10 °C warmer than those of western Canada. The effect of cold condensation on the fate of organochlorine in these areas remains unknown, but is likely to result in more pronounced accumulation of toxic compounds than we have observed in our study area. □

Received 9 January; accepted 21 July 1998.

- Donald, D. B. *et al.* Polychlorinated Biphenyls and Organochlorine Pesticides in the Aquatic Environment along the Continental Divide Region of Alberta and British Columbia (Spec. Rep. Inland Waters Directorate, Environment Canada, Regina, Saskatchewan, 1993).
- Welch, H. E. *et al.* Brown snow: a long-range transport event in the Canadian Arctic. *Environ. Sci. Technol.* **25**, 280–286 (1991).
- Wania, F. & Mackay, D. Global fractionation and cold condensation of low volatility organochlorine compounds in polar regions. *Ambio* **22**, 10–18 (1993).
- Simonich, S. L. & Hites, R. A. Global distribution of persistent organochlorine compounds. *Science* **269**, 1851–1854 (1995).
- Miskimmin, B. *et al.* Chlorobornanes in sediments and fish thirty years after toxaphen treatment of lakes. *Environ. Sci. Technol.* **29**, 2490–2495 (1995).
- Campbell, L. M. *The Use of Stable Isotope Ratios to Discern Organochlorine Bioaccumulation Patterns in a Sub-alpine Rocky Mountain Lake Food Web*. Thesis, Univ. Alberta (1997).
- Kidd, K. A. *et al.* High toxaphene concentrations in fish from a subarctic lake. *Science* **269**, 240–242 (1995).
- Kidd, K. A. *et al.* Correlation between stable nitrogen isotope ratios and concentrations of organochlorines in biota from a freshwater food web. *Sci. Tot. Environ.* **160/161**, 381–390 (1995).
- Weathers, K. C. *et al.* A regional acidic/fog water event in the eastern United States. *Nature* **319**, 657–658 (1986).
- Baltensperger, U. *et al.* Transfer of atmospheric constituents into an alpine snow field. *Atmos. Environ.* **A 27**, 1881–1890 (1993).
- Muir, D. C. G. *et al.* Organochlorine contaminants in arctic marine food chains: accumulation of specific polychlorinated biphenyls and chlordane-related compounds. *Environ. Sci. Technol.* **22**, 1071–1079 (1988).
- Muir, D. C. G. *et al.* Spatial trends and historical profiles of organochlorine pesticides in Arctic lake sediments. *Sci. Tot. Environ.* **160/161**, 447–457 (1995).
- Hinckley, D. A. *et al.* Determination of vapor pressures for nonpolar and semipolar organic compounds from gas chromatographic retention data. *J. Chem. Eng. Data* **35**, 323–327 (1990).
- Muir, D. C. G. *et al.* Spatial trends and historical deposition of polychlorinated biphenyls in Canadian mid-latitude and arctic lake sediments. *Environ. Sci. Technol.* **30**, 3609–3617 (1996).
- Muir, D. C. G. *et al.* Geographic variation of chlorinated hydrocarbons in burbot (*Lota lota*) from remote lakes and rivers in Canada. *Arch. Environ. Contam. Toxicol.* **19**, 530–542 (1990).
- Gregor, D. J. in *Seasonal Snowpacks* (ed. Davies, T.) 323–358 (NATO Ser. G, Vol. 28, Springer, Berlin, 1991).
- Wania, F. Modelling the fate of non-polar organic chemicals in an aging snowpack. *Chemosphere* **35**, 2345–2363 (1997).

- Metcalf, R. L. *Organic Insecticides, their Chemistry and Mode of Action* (Interscience, New York, 1955).
- Canadian Arctic Contaminant Assessment Report (Minister of Public Works and Government Services, Ottawa, Canada, 1997).
- Schindler, D. W. & Pacas, C. in *Ecological Outlooks Project* (eds Green, J., Pacas, C., Bayley, S. & Cornwall, L.) (Banff-Bow Valley Study, Dept of Canadian Heritage, Ottawa, 1996).
- Mackay, D., Shiu, W. Y. & Ma, K.-C. *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals* Vol. V (Lewis Publishers, Boca Raton, FL, 1997).
- Mackay, D., Shiu, W. Y. & Ma, K.-C. *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals* Vol. 1. *Monoaromatic Hydrocarbons Chlorobenzenes and PCBs* (Lewis Publishers, Chelsea, Michigan, 1992).

**Acknowledgements.** We are grateful to the Banff Warden Office, in particular D. Gilbride, for help with locating weather stations and providing meteorological data. We thank the ski patrol crews from Lake Louise, Sunshine, Tod and Whistler ski mountains for their support and interest. Sample fractionation and GC analyses were conducted at the Freshwater Institute, Winnipeg. R. Crosley and L. Mottle from Environment Canada, Calgary, supplied expertise, sampling equipment and laboratory facilities for this work. This work was supported by the Canadian Chlorine Coordinating Committee (C4), an NSERC Strategic Grant, and an NSERC Operating Grant (D.W.S.).

Correspondence and requests for materials should be addressed to J.M.B. at the University of Ottawa (e-mail: blais@science.uottawa.ca).

## The influence of natural mineral coatings on feldspar weathering

M. A. Nugent<sup>†</sup>, S. L. Brantley<sup>\*</sup>, C. G. Pantano<sup>‡</sup> & P. A. Maurice<sup>§</sup>

<sup>\*</sup> Department of Geosciences, <sup>‡</sup> Department of Material Science and Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, USA  
<sup>§</sup> Department of Geology, Kent State University, Kent, Ohio 44242, USA

Quantification of the rate of weathering of feldspar, the most abundant mineral in the Earth's crust, is required to estimate accurately carbon dioxide fluxes over geological timescales and to model groundwater chemistry. Laboratory dissolution rates, however, are consistently found to be up to four orders of magnitude higher than the 'natural' rates<sup>1,2</sup> measured in the field. Although this discrepancy has been attributed to several factors<sup>3</sup>, previous research has tended to suggest that the underlying mechanism of feldspar dissolution under acidic pH may differ between the field and the laboratory<sup>3</sup>. Here we demonstrate that weathered albite surfaces, like laboratory-dissolved samples, are sodium- and aluminium-depleted, indicating that the dissolution mechanism in acidic soils is similar to that in acidic laboratory solutions. We find that microtopography images are consistent with dissolution occurring at specific surface sites—indicative of surface-controlled dissolution dominated by a non-stoichiometric layer. Elevated aluminium and silicon ratios reported previously<sup>3,4</sup>, and used to suggest a mechanism for field weathering different from laboratory dissolution<sup>3</sup>, can alternatively be explained by a thin, hydrous, patchy, natural coating of amorphous and crystalline aluminosilicate. This coating, which is largely undetected under scanning electron microscopy after cleaning, but visible under atomic force microscopy, alters surface chemistry measurements and may partially inhibit the field dissolution rate.

Numerous laboratory dissolution studies on feldspar at acidic pH reveal Na/Ca/K- and Al-depleted surfaces, as documented by X-ray photoelectron spectroscopy (XPS)<sup>5–10</sup>, secondary ion mass spectrometry (SIMS)<sup>6,8,9,11,12</sup> and other techniques<sup>13–18</sup>. Depleted surfaces have been used to suggest that the mechanism of feldspar dissolution in the laboratory involves detachment of an Al-depleted, H-enriched surface complex<sup>14,17,18</sup>. In contrast, Al depletion on naturally weathered feldspar has never been reported. Instead, either elevated Al/Si ratios<sup>3,4</sup> or Al/Si ratios unaltered from the

<sup>†</sup> Present address: Department of Geosciences, SUNY at Stony Brook, Stony Brook, New York 11794-2100, USA.

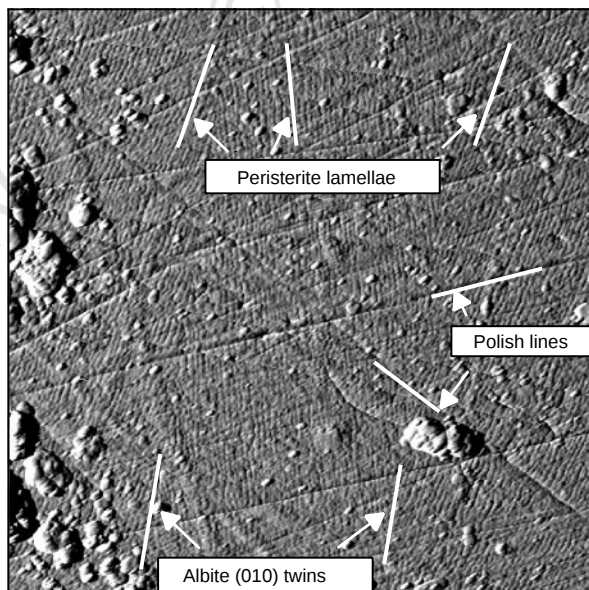


bulk<sup>19</sup> are reported for feldspars from acidic soils. Nesbitt and Muir<sup>3</sup> inter that the surface Al/Si enrichment of oligoclase removed from a 10,000-year-old Canadian (acidic) spodosol is due to Si depletion, which implies that the dissolution mechanism differs from that in the laboratory.

To assess the mechanism of feldspar dissolution in an acidic soil, albite tablets ( $\text{Ab}_{90}\text{An}_{10}$ , which is 90 mol % albite, 10 mol % anorthite) from Havre de Grace, Quebec, were polished to 0.25  $\mu\text{m}$  on the (001) cleavage face and buried in the B horizon of a spodosol (State College, Pennsylvania). The 'albite' samples are peristerites (alternating albite and oligoclase lamellae) with a bulk composition within the albite compositional range. Several holey bottles containing albite tablets and soil were buried in August 1993, and one bottle was removed after 0.5, 1, 2, 3 and 3.5 years. Upon retrieval, the 6-month, 1- and 2-year samples were briefly rinsed in water then ultrasonicated in acetone. The 3- and 3.5-year tablets were ultrasonicated in high-purity acetone for 20 and 45 min, respectively, and stored in a desiccator. Polished unweathered blanks were prepared and left unweathered in the desiccator.

The burial site is located in a humid, temperate climate, which is vegetated by conifers and receives 103 cm of acidic (pH 4.3) precipitation annually. The B horizon (quartz, minor kaolinite and minor iron oxides) is overlain by a thin (<5 cm), organic-rich A horizon. Soil pore waters, sampled at burial depth by suction lysimeter in 1995–96 varied in pH from 4.9 to 5.6 (average, 5.1) and were under- and supersaturated with respect to all feldspars and kaolinite, respectively.

Surface microtopography (Fig. 1), which is not present before burial, indicates that dissolution occurred during weathering. Etching of specific surface sites, as observed under Tapping Mode atomic force microscopy (AFM) is consistent with a surface-controlled dissolution mechanism, in agreement with previous scanning electron microscope (SEM) observations for naturally weathered and laboratory-dissolved feldspars<sup>20–22</sup>. Based on microtopography development over 3 years, our field dissolution rate<sup>23</sup>, estimated between  $10^{-16}$  and  $10^{-18}$  mol<sub>feldspar</sub> cm<sup>-2</sup> s<sup>-1</sup> is comparable



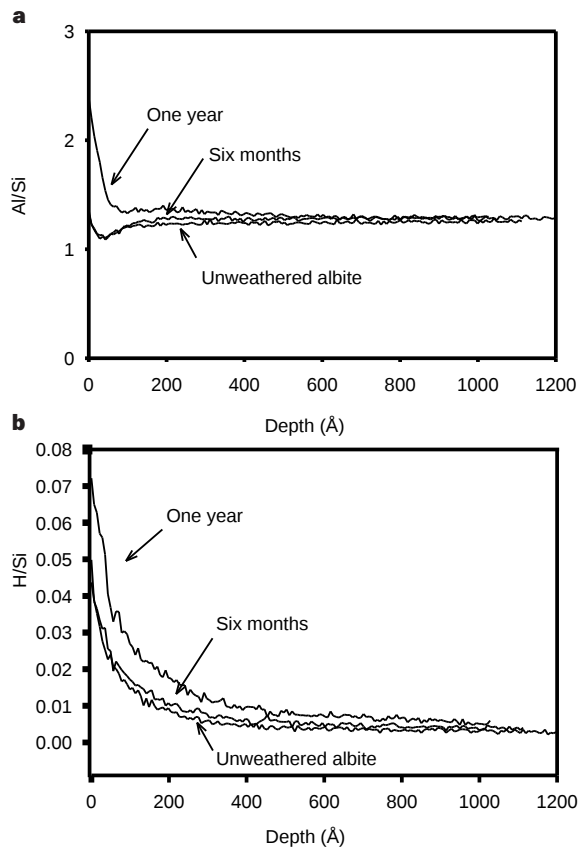
**Figure 1** Tapping Mode instrument AFM amplitude image ( $2.5 \times 2.5 \mu\text{m}$ ) in a coating-free area of peristeritic albite weathered for 3 years. The microtopography is the result of preferential etching of albite (010) twins and albite and oligoclase lamellae, as confirmed by TEM analysis. This microtopography, absent on unweathered samples and never observed under SEM, is present in patches on the sample weathered for 6 months, and everywhere imaged on the 3- and 3.5-year samples. The grooves created by the peristerite lamellae are at least 1.5 nm deep, on average<sup>21</sup>, and polish lines are present before and after weathering.

to other reported field rates<sup>2</sup>, suggesting that our samples dissolved similarly to feldspars weathering naturally.

Using SIMS profiling, no difference in the Al/Si ratio is observed between the unweathered and 6-month samples, but the Al/Si ratio is increased at least 500 Å into the feldspar surface weathered for 1 year (Fig. 2a). Our 1-year profiles are similar to SIMS profiles<sup>3</sup> for oligoclase naturally weathered in an acidic spodosol. XPS analyses (Fig. 3) reveal that the 6-month sample is Al-depleted, but that samples weathered longer than 6 months are Al-enriched. Every weathered albite surface is Na-deficient.

We find that a patchy coating, consisting of particles ranging in size from nanometres to millimetres and in thickness from ~5 nm to 1  $\mu\text{m}$ , covers at least 35% of the 3-year sample after cleaning, as imaged under AFM. This coating, largely undetected under SEM but obvious under AFM, is sufficiently thick to affect XPS (upper ~80 Å) and SIMS (depth profiling) signals. The Na- and Al-depleted 6-month sample has significantly less coating, as observed under AFM.

Sufficient coating coverage will result in higher Al/Si and lower Na/Si ratios than those for unweathered feldspar. The coating, analysed by energy dispersive X-ray analysis (EDX) during transmission electron microscopy (TEM), contains crystalline kaolinite, which has a higher Al/Si and lower Na/Si ratio than albite, in discrete patches within the otherwise amorphous coating material. Amorphous portions of the coating are hydrous (Fig. 2b) and contain Al, Si, O and minor Fe. Because the 6-month sample shows Al and Na depletion and has less coating material, we argue



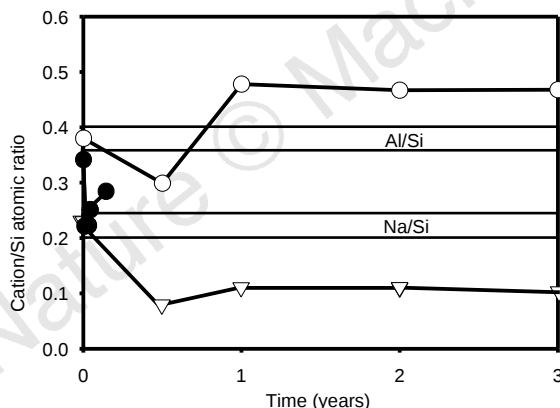
**Figure 2.** SIMS signal ratios as a function of depth into the albite surface. **a**, Al/Si; **b**, H/Si. After gold-coating the samples, a gold-coated TEM grid was laid over the area to be analysed, to provide sufficient charge stabilization during analysis. SIMS profiles were then collected on a Cameca Model IMS/3F using a 50 nA, 15.4 eV, negative O-18 primary beam. The ~75- $\mu\text{m}$  beam was rastered over a  $250 \times 250 \mu\text{m}$  area, and positive secondary ions were collected from an area in the centre of the crater which was ~10  $\mu\text{m}$  in diameter. Because AFM analyses reveal that the coating is patchy, coating thickness cannot accurately be assessed by only our SIMS data.

that the albite surface during weathering, like laboratory-dissolved surfaces, is Al- and Na-depleted. A natural coating obscures the feldspar surface XPS and SIMS signals on our other field-weathered samples, because these signals average the Al and Si compositions over coated and uncoated areas of the surface.

A 3.5-year sample, ultrasonicated for 45 min within 1 day of retrieval, revealed relatively low Al/Si (0.32) and Na/Si (0.11) atomic ratios under XPS. Under AFM, the coating coverage on this 3.5-year sample was significantly less than on the 3-year sample. The Na- and Al-depleted signal on this 3.5-year sample, which presumably includes areas covered by coating during weathering, suggests the feldspar surface is Na- and Al-depleted underneath the coating during weathering.

The 3-year sample was ultrasonicated for an additional 20 min, 9 months after retrieval, and re-analysed by XPS. The Al/Si atomic ratio remained high (0.47), the Na/Si ratio remained low (0.08), and the coating coverage under AFM remained approximately the same. The difficulty in removing the coating from the 3-year sample several months after retrieval, and the ease with which coating is removed from the 3.5-year sample immediately after recovery, is consistent with ageing into an adherent layer.

A second set of five samples, buried later and recovered after 1, 2 (2 samples), 3, and 4 months, reveal variable but elevated Al/Si atomic ratios (0.40, 0.40 and 0.51, 0.38, and 0.48, respectively) under XPS, demonstrating that the coating is established very early in weathering. Furthermore, coating removal is difficult and not necessarily reproducible. For example, the two 2-month samples, which were cleaned identically, have Al/Si ratios of 0.40 and 0.51. Coating adherence is extremely sensitive to the timing and method of cleaning, and possibly, therefore, to wet/dry cycles in soils and other seasonal and environmental factors. Finally, because even prolonged cleaning might not remove such a coating from natural



**Figure 3** Al/Si (empty circles) and Na/Si (empty triangles) atomic ratios for albite versus time of weathering, as measured by XPS. The boxes are defined by the high and low values of 11 Al/Si and Na/Si ratios measured on polished, unweathered albite blanks. We note that cleaning unweathered samples with and without a rinsing in water during polishing resulted in indistinguishable Al/Si values. All samples were placed in a ultraviolet ozone cleaning chamber<sup>28</sup> for ~30 min to remove adventitious carbon, then Al 2p, Si 2p and Na KLL peaks were measured using a Kratos Analytical XPS 800pci instrument. The atomic % Al ranges from 7.1 to 7.9 for unweathered samples, and is 4.2, 7.6, 7.2 and 9.5, for samples weathered for 6 months, 1, 2 and 3 years, respectively. The atomic % Na ranges from 3.9 to 5.2 for unweathered samples, but is never higher than 2.6 during weathering. The atomic % Si ranges from 18.1 to 21.6 in unweathered samples. After weathering, the atomic % Si is 26.4, 17.5, 17.3 and 20.2 at 6 months, 1, 2 and 3 years, respectively. For comparison, XPS Al/Si atomic ratios<sup>10</sup> (filled circles) for Amelia Courthouse albite treated with pH 4 acetate/lithium-acetate buffer (25°C) laboratory solutions<sup>29</sup> are plotted. The rise in Al/Si with the duration of dissolution might be attributed to a back-reaction involving the development of an Al-rich surface coating similar to the one we observe on our naturally weathered samples.

samples, we suggest that a natural coating may explain previous observations<sup>3,4</sup> of Al enrichment on naturally weathered feldspars.

A simple mass balance based on minimum coating thickness and coverage and the minimum feldspar dissolution rate<sup>24</sup> shows that the coating contains more Al and Si than is liberated by feldspar dissolution, suggesting that the coating is at least partially derived from the soil pore waters. This argument is further strengthened by considering that the pore waters are supersaturated with respect to kaolinite (found in the coating) and that indigenous soil quartz grains in this feldspar-free soil also have an Al-, Si-rich coating, based on SEM-EDX.

Casey and others<sup>25</sup> have argued that the formation of secondary precipitates during weathering occurs by epitaxial growth on an Al-depleted, Si-enriched surface on silicate minerals. Our observations document for the first time, to our knowledge, that such a depleted layer may develop on naturally weathered samples. However, the lack of strong coating adherence before ageing further supports our suggestion that secondary phase formation occurs, in part, as silicic acid and aluminium ions precipitate from solution. Furthermore, our observation of an Al-depleted layer does not necessarily prove that this layer is needed for secondary mineral precipitation; indeed, our observations of aluminosilicate coatings on soil quartz grains suggests that epitaxial growth is not required for secondary precipitate formation.

On the basis of the development of dissolution textures (Fig. 1), feldspar dissolution occurs despite the presence of a patchy coating, and presumably, through the coating. For example, the 3.5-year sample, which probably had at least as much coating coverage as the 3-year sample (~35%) while buried, showed <35% coverage under AFM after cleaning for 45 min. Yet, because both the 3- and 3.5-year samples showed dissolution texture everywhere, we infer some of these textured areas dissolved despite the presence of an overlying coating.

The coating may, however, partially inhibit dissolution, thus contributing to the laboratory/field-rate discrepancy. For example, stagnant pore waters in a feldspar-porewater-coating 'sandwich' may be more supersaturated with respect to albite than bulk soil pore water<sup>26</sup>. The lowered driving force for dissolution in such a microenvironment may partially explain the laboratory/field discrepancy, even if the  $pH_{\text{micropore}} < pH_{\text{porewater}}$ , as we might predict on the basis on experiments in which Al-rich coatings are precipitated<sup>27</sup>. We conclude that the laboratory and field dissolution mechanisms for feldspar are similar during early weathering, but that coating precipitation reactions may affect feldspar dissolution over geological timescales. □

Received 14 January; accepted 27 July 1998.

- Blum, A. & Stillings, L. in *Chemical Weathering Rates of Silicate Minerals* 291–352 (Mineralogical Society of America, Washington DC, 1995).
- White, A. in *Chemical Weathering Rates of Silicate Minerals* 407–462 (Mineralogical Society of America, Washington DC, 1995).
- Nesbitt, H. W. & Muir, I. SIMS depth profiles of weathered plagioclase and processes affecting dissolved Al and Si in some acidic soil solutions. *Nature* **334**, 336–338 (1988).
- Blum, A. E., White, A. F. & Hochella, M. F. The surface chemistry and topography of weathered feldspar surfaces. *Trans. Am. Geophys. Union, Abstr. Prog.* **73**, 602 (1992).
- Casey, W., Westrich, H. R. & Arnold, G. W. The surface chemistry of labradorite feldspar reacted with aqueous solutions at pH = 2, 3, and 12. *Geochim. Cosmochim. Acta* **52**, 2795–2807 (1988).
- Goossens, D. et al. A SIMS, XPS, SEM, TEM and FTIR study of feldspar surfaces after reacting with acidic solutions. *Proc. 6th Int. Symp. of Water-Rock Interaction* (ed. Miles, D. L.) 267–270 (Balkema, Rotterdam, 1989).
- Hellman, R., Eggleston, C. R., Hochella, M. F. & Crerar, D. A. The formation of leached layers on albite surfaces during dissolution under hydrothermal conditions. *Geochim. Cosmochim. Acta* **54**, 1267–1281 (1990).
- Muir, I., Bancroft, G. R. & Nesbitt, H. W. Characteristics of altered labradorite surfaces by SIMS and XPS. *Geochim. Cosmochim. Acta* **53**, 1235–1241 (1989).
- Muir, I., Bancroft, G. M., Shetye, W. & Nesbitt, H. W. A SIMS and XPS study of dissolving plagioclase. *Geochim. Cosmochim. Acta* **54**, 2247–2256 (1990).
- Rufe, E. *Influence of Surface Chemistry on the Natural and Artificial Weathering of Feldspars in the Soil Environment*. Thesis, Penn State Univ. (1994).
- Muir, I. & Nesbitt, H. W. Effects of aqueous cations on the dissolution of labradorite feldspar. *Geochim. Cosmochim. Acta* **55**, 3181–3189 (1991).
- Muir, I. & Nesbitt, H. W. Controls on differential leaching of calcium and aluminum from labradorite in dilute electrolyte solutions. *Geochim. Cosmochim. Acta* **56**, 3979–3985 (1992).
- Hochella, M. F., Ponader, H. B., Turner, A. M. & Harris, D. W. The complexity of mineral dissolution as viewed by high resolution scanning Auger microscopy: Labradorite under hydrothermal conditions. *Geochim. Cosmochim. Acta* **52**, 385–394 (1988).

14. Hellmann, R., Dran, J.-C. & Della Mea, G. The albite-water system: Part III. Characterization of leached and hydrogen-enriched layers formed at 300 °C using MeV ion beam techniques. *Geochim. Cosmochim. Acta* **61**, 1575–1594 (1997).
15. Casey, W., Westrich, H. R., Arnold, G. W. & Banfield, J. R. The surface chemistry of dissolving labradorite feldspar. *Geochim. Cosmochim. Acta* **58**, 821–832 (1994).
16. Chou, L. & Wollast, R. Steady state kinetics and dissolution mechanism of albite. *Am. J. Sci.* **285**, 963–993 (1985).
17. Brantley, S. & Stillings, L. Feldspar dissolution at 25 °C and low pH. *Am. J. Sci.* **296**, 101–127 (1996).
18. Blum, A. & Lasaga, A. Role of surface speciation in the low-temperature dissolution of minerals. *Nature* **331**, 431–433 (1988).
19. Lasaga, A. C. in *Chemical Weathering Rates of Silicate Minerals* 23–86 (Mineralogical Society of America, Washington DC, 1995).
20. Berner, R. A. & Holdren, G. R. Mechanism of feldspar weathering—II. Observations of feldspars from soils. *Geochim. Cosmochim. Acta* **43**, 1173–1186 (1979).
21. Holdren, G. R. & Berner, R. A. Mechanism of feldspar weathering—I. Experimental studies. *Geochim. Cosmochim. Acta* **43**, 1161–1171 (1979).
22. Lee, M. & Parsons, I. Microtextural controls of weathering of perthitic alkali feldspars. *Geochim. Cosmochim. Acta* **59**, 4465–4488 (1995).
23. Nugent, M. A., Maurice, P. A. & Brantley, S. L. in *Proc. 9th Int. Symp. of Water–Rock Interaction* (eds Arehart, G. B. & Hulston, J. R.) 387–390 (Balkema, Rotterdam, 1998).
24. Nugent, M. *Temporal Evolution of Feldspar Surfaces during the Initial Stages of in-situ Weathering*. Thesis, Penn State Univ. (1997).
25. Casey, W. H., Benfield, J. F., Westrich, H. R. & McLaughlin, L. What do dissolution experiments tell us about natural weathering? *Chem. Geol.* **105**, 1–15 (1993).
26. Hochella, M. F. & Banfield, J. F. in *Chemical Weathering Rates of Silicate Minerals* 353–406 (Mineralogical Society of America, Washington DC, 1995).
27. Violante, A., Gianfreda, L. & Violante, P. Effect of prolonged aging on the transformation of short-range ordered aluminum precipitation products formed in the presence of organic and inorganic ligands. *Clays Clay Miner.* **41**, 353–359 (1993).
28. Briggs, D. & Seah, M. P. *Practical Surface Analyses Vol. 1, Auger and X-ray Photoelectron Spectroscopy* (Wiley, New York, 1990).
29. Murphy, K. M. *Kinetics of Albite Dissolution: the Effect of Grain Size*. Thesis, Univ. Wyoming (1993).

**Acknowledgements.** We thank V. Bojan, D. Voigt, S. Yau, E. Rufe and B. Turner for discussions, M. Lee for feldspar TEM analyses, J. Hamilton for XPS analyses and Y. Chen and C. Doud for assistance with burial experiments. This work was funded by a grant from the Office of Basic Energy Sciences, Dept of Energy.

Correspondence and requests for materials should be addressed to M.A.N. (e-mail: nugent@sbmp04.ess.sunysb.edu).

## Evidence from the rare-earth-element record of mantle melting for cooling of the Tertiary Iceland plume

C. Tegner<sup>††</sup>, C. E. Leshner<sup>‡‡</sup>, L. M. Larsen<sup>\*§</sup> & W. S. Watt<sup>\*§</sup>

<sup>\*</sup> Danish Lithosphere Centre, Øster Voldgade 10, DK-1350 København K, Denmark

<sup>††</sup> Department of Earth Sciences, University of Aarhus, DK-8000 Aarhus C, Denmark

<sup>‡‡</sup> Department of Geology, University of California, Davis, California 95616, USA

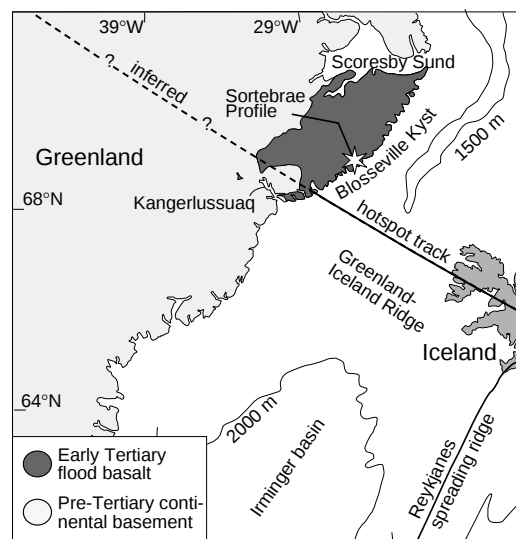
<sup>§</sup> Geological Survey of Denmark and Greenland, Thoravej 8, DK-2400 København NV, Denmark

Widespread flood basalt volcanism and continental rifting in the northeast Atlantic in the early Tertiary period (~55 Myr ago) have been linked to the mantle plume now residing beneath Iceland<sup>1–5</sup>. Although much is known about the present-day Iceland plume<sup>6–9</sup>, its thermal structure, composition and position in the early Tertiary period remain unresolved. Estimates of its temperature, for example, range from >1,600 °C in some plume models<sup>3</sup> to ~1,500 °C based on the volume and composition of basaltic crust<sup>10–12</sup>. Several recent studies<sup>4</sup> have emphasized similarities in the thermal and chemical structure of the Tertiary and present-day plumes to argue for stability of the mantle anomaly, whereas others<sup>12,13</sup> relate variations in basalt volumes and compositions to changes in plume flux. Moreover, some authors<sup>12,13</sup> have assumed that the plume was rift-centred for its entire history, whereas others argue that it became ridge-centred only after plate separation<sup>14,15</sup>. Here we report compositional data for ~6,000 metres of flood basalts erupted in east Greenland, close to the inferred plume axis, that we use to constrain the Tertiary plume structure. Rare-earth-element systematics place limits on the

pressures and extents of mantle melting and show that the mantle was initially moderately hot (~1,500 °C), but that its temperature declined during flood volcanism. These observations are difficult to reconcile with current plume-head models, and call for important lithospheric control<sup>5,10,16–18</sup> on actively upwelling mantle along the rifted margin.

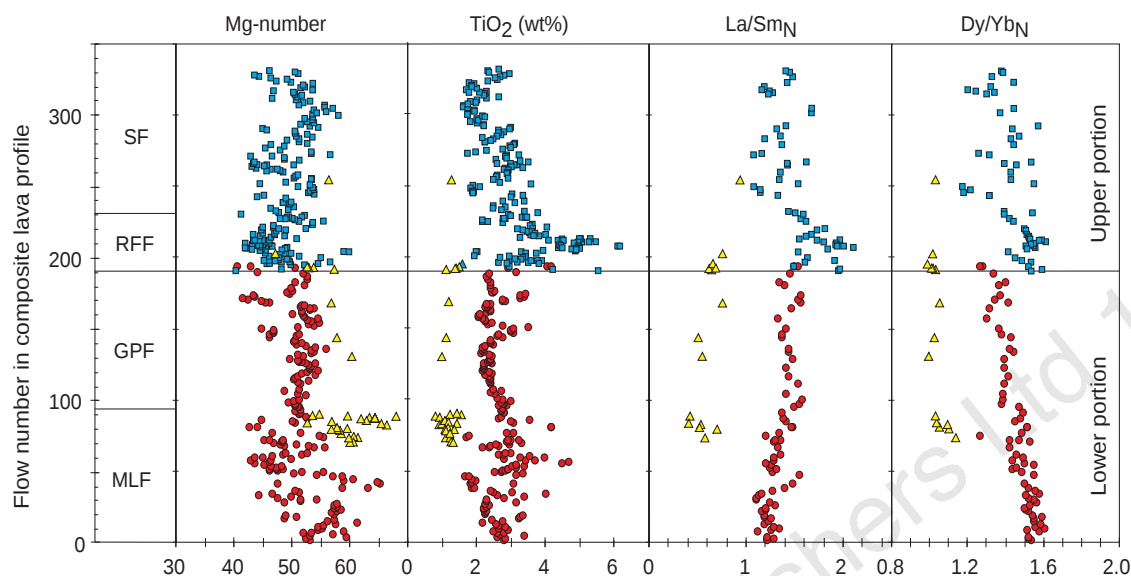
The east Greenland flood-basalt succession was emplaced in less than 1–2 Myr in the early Tertiary period<sup>19</sup>, and thickens southwards to >6 km towards the supposed hotspot track (Greenland–Iceland–Faeroes ridge; Fig. 1)<sup>20</sup>. Data on major- and rare-earth elements (REE) for a composite profile of the lava succession in the coastal mountains of the Sortebrae area (Fig. 1), totalling 331 flow units in four mappable formations<sup>20</sup>, are given in Fig. 2 (and in table form in Supplementary Information). Geochemically, these formations are composed of two distinct, uncontaminated (see Methods) suites; iron- and titanium-rich lavas (1.63–6.18 wt% TiO<sub>2</sub>) are found throughout, and low-Ti lavas (0.82–1.96 wt% TiO<sub>2</sub>) are restricted to the middle portion of the succession. The high-Ti suite has elevated chondrite-normalized<sup>21</sup> La/Sm<sub>N</sub> (1.09–2.11) and Dy/Yb<sub>N</sub> (1.18–1.62), while the low-Ti suite has near-chondritic to sub-chondritic REE ratios (La/Sm<sub>N</sub> = 0.42–0.94; Dy/Yb<sub>N</sub> = 0.99–1.14). Preliminary Sr–Nd isotope data for a subset of lavas confirm the highly depleted nature of the low-Ti suite and source (measured  $\epsilon_{Nd}$  = +8 to +11;  $^{87}Sr/^{86}Sr$  = 0.7026–0.7029) compared to that for the high-Ti suite ( $\epsilon_{Nd}$  = +5 to +8;  $^{87}Sr/^{86}Sr$  = 0.7030–0.7034; O. Stecher, unpublished data). In terms of major elements, REE and Sr–Nd isotopes, the low-Ti suite closely resembles depleted North Atlantic mid-ocean-ridge basalts (MORBs) formed by large-degree mantle melting (~15–25%; refs 22, 23) (Fig. 3a).

The systematic variations in REE ratios shown in Fig. 2 are most striking. In the lower portion of the succession (flow numbers 1–193; coded red circles) La/Sm<sub>N</sub> for the high-Ti suite increases up-section from ~1.2 to ~1.6, whereas the Dy/Yb<sub>N</sub> ratio decreases from ~1.6 to ~1.3 and correlates inversely with La/Sm<sub>N</sub> (Figs 2 and 3a). For the same stratigraphic interval, the variations in La/Sm<sub>N</sub> and Dy/Yb<sub>N</sub> for the low-Ti suite (yellow triangles) roughly parallel the high-Ti suite. In contrast, La/Sm<sub>N</sub> and Dy/Yb<sub>N</sub> for the high-Ti



**Figure 1** Map of northeast Atlantic. Map showing exposed Tertiary flood basalt along the east Greenland volcanic rifted margin (~55 Myr old) and in Iceland (<16 Myr old), and ocean floor bathymetry. The solid line shows the track of the Iceland hotspot after continental breakup and the dashed line shows the inferred continuation of the plume track beneath Greenland<sup>14</sup>.





**Figure 2** Chemical stratigraphy. Chemical stratigraphy for the ~6-km-thick composite profile of the east Greenland flood basalt succession (331 flow units) sampled in the Sortebræe area, Blosseville Kyst; location is shown in Fig. 1. From left to right, the diagram shows the subdivision into four mappable formations (Milne Land Formation (MLF), Geikie Plateau Formation (GPF), Rømer Fjord Formation (RFF), and Skraenterne Formation (SF)); whole-rock Mg-number (atomic  $100 \text{ Mg}/(\text{Mg} + \text{Fe}^{2+})$ , assuming  $\text{Fe}_2\text{O}_3/\text{FeO} = 0.15$ );  $\text{TiO}_2$  content (wt%);  $\text{La}/\text{Sm}_N$  and  $\text{Dy}/\text{Yb}_N$  (see also Supplementary Information). The low-Ti suite is shown as yellow triangles. The high-Ti suite is shown as red circles for the lower portion (MLF and GPF) and blue squares for the upper portion (RFF and SF) of the succession.

suite in the upper portion of the succession (flow numbers 190–331; coded blue squares) are highly variable and correlate positively (Figs 2 and 3a). These systematics, coupled with a lack of correlation with Mg-number and  $\text{TiO}_2$  content (Fig. 2), suggest that the REE compositions are related to mantle melting processes and source differences, and are only marginally affected by crustal contamination and crystal fractionation.

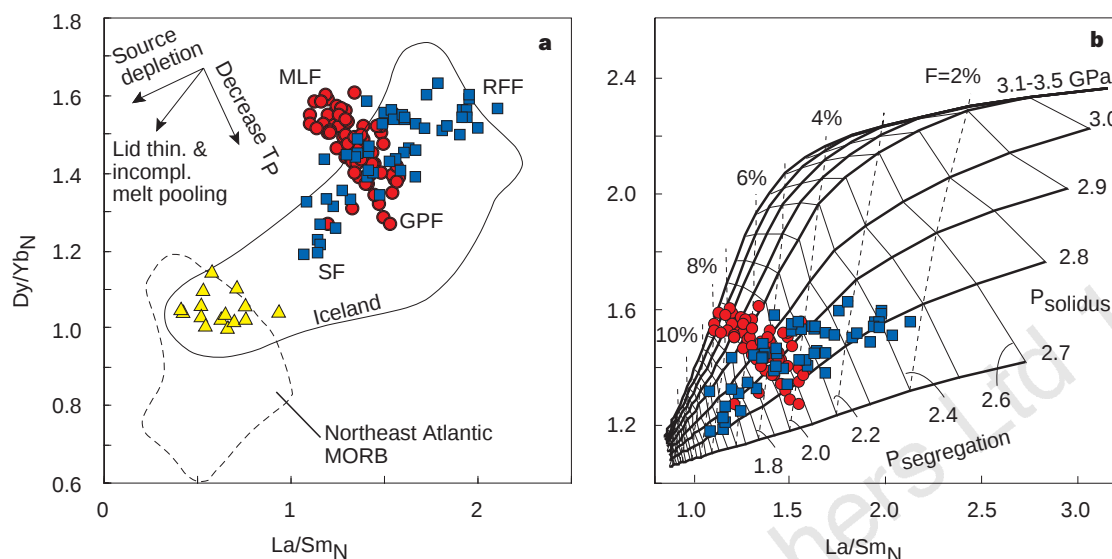
Given that Sm is less incompatible than La during peridotite melting, the  $\text{La}/\text{Sm}_N$  ratio decreases as the extent of melting increases. Yb is significantly more compatible than Dy in garnet so the  $\text{Dy}/\text{Yb}_N$  ratio can be related to the proportion of melt derived from garnet-bearing sources<sup>10</sup>. If the increase in  $\text{La}/\text{Sm}_N$  reflects a decline in the mean extent of melting ( $F$ ), and the decline in  $\text{Dy}/\text{Yb}_N$  corresponds to lower mean pressures of melting ( $P$ ), then the antithetical relationship for the high-Ti suite of the Milne Land Formation (MLF) and Geikie Plateau Formation (GPF) (Fig. 3a) implies a temporal decline in both  $F$  and  $P$ . This is readily explained by decreasing mantle potential temperature<sup>23</sup>. A similar relationship between  $F$  and  $P$  may be explained by progressive dehydration of metasomatically enriched mantle, but this is considered unlikely given the tholeiitic character of primitive, uncontaminated high- and low-Ti basalts, and the observation that the earliest erupted lavas of each suite are among the most light-REE depleted compositions. Neither changes in mantle temperature nor variations in volatile components account for the positive correlation between  $\text{La}/\text{Sm}_N$  and  $\text{Dy}/\text{Yb}_N$  shown for the Rømer Fjord Formation (RFF) and the Skraenterne Formation (SF) (Fig. 3a), implying an inverse relationship between  $F$  and  $P$ . Such a relationship between these melting parameters is consistent with decompression melting controlled by variations in lithosphere thickness<sup>10</sup>. However, the up-section fluctuations in  $\text{La}/\text{Sm}_N$  and  $\text{Dy}/\text{Yb}_N$  suggest a process more closely linked to the dynamics of melt segregation, as documented in Iceland<sup>7</sup>.

Mantle conditions for the generation of primary melts for the high-Ti suite can be quantified by forward modelling of polybaric melting. Such modelling depends on melting rate, reaction stoichiometry and partition coefficients for which we use available experimental data, and on source composition for which we assume a slightly depleted anhydrous mantle analogous to that proposed by

Fram and Leshner<sup>24</sup> for the earlier east Greenland Tertiary flood basalts (see Methods). Heavy curves in Fig. 3b show the pooled melt compositions predicted for decompression melting, contoured for final segregation pressure ( $P_{\text{segregation}}$  is at the top of melting column) and mean melt fraction ( $F$ ) as light and dashed curves, respectively.

The model decompression paths shown in Fig. 3b are roughly orthogonal to the elongated array defined by the high-Ti suite in the lower part of the succession (MLF and GPF), and subparallel to the array for the upper part (RFF and SF). For example, the MLF and GPF basalts form an array bounded by  $P_{\text{segregation}}$  contours for 1.9 and 2.2 GPa and decompression melting paths initiating at solidus pressures ( $P_{\text{solidus}}$ ) of 3.5 and 2.7 GPa. If a high (dry) mantle solidus is assumed<sup>25</sup>,  $P_{\text{solidus}}$  of 3.5 GPa implies a maximum potential temperature ( $T_P$ ) of ~1,500 °C for the generation of the basal lavas of the MLF. The up-section decline in  $P_{\text{solidus}}$  to 2.7 GPa implies a drop in  $T_P$  of ~80 °C. These  $T_P$  estimates (1,500–1,420 °C) fall within the range proposed for the modern Iceland plume ( $\geq 1,500$  °C; refs 9, 26, 27) and for modern ambient mantle (1,280–1,310 °C; refs 11, 25), and are slightly below the inferred potential temperatures based on early Tertiary igneous crustal thickness in the northeast Atlantic and REE inversion methods (~1,500 °C; refs 11, 12). Mean melt fractions ( $F$ ) of ~8–4% for the lower portion of the east Greenland volcanic succession are less than those estimated for lavas comprising the seaward-dipping reflector sequence elsewhere along the volcanic rifted margin (~15%; ref. 28) and for modern Iceland (~20%; ref. 23), but are consistent with limitations in mantle upwelling imposed by the continental lithosphere at the initiation of rifting<sup>10</sup>.

The data in Fig. 3b suggest that the basalts of the RFF and SF formed under relatively uniform  $P_{\text{solidus}}$  (~2.9 GPa), and therefore at near-constant  $T_P$  (~1,440 °C), but segregated from a wide range of pressures (1.4–2.5 GPa), similar to Icelandic basalts<sup>7</sup>. There is a weak indication of an up-section decline in  $P_{\text{segregation}}$ , which may reflect a shoaling of the melting column related to plate separation. It is also clear that the lowest- $F$  melts with the highest titanium contents recovered from the base of the RFF, and the scarcity of low-Ti basalts above this level, represent a fundamental change in the conditions of melt generation. We speculate that initial mantle upwelling beneath partially extended continental lithosphere



**Figure 3** Mantle melting systematics. **a**, Covariation of  $\text{La}/\text{Sm}_N$  and  $\text{Dy}/\text{Yb}_N$  for the lavas shown in Fig. 2. Also shown are the fields contouring the composition of Icelandic basalts<sup>8</sup> and “normal” mid-ocean ridge basalt (MORB) generated from ambient northeast Atlantic mantle outside the compositional influence of the Iceland plume<sup>22</sup>. Symbols as in Fig. 2. **b**, Data fields from **a** superimposed on model curves for pooled mantle melts (see Methods). Heavy solid curves show

resulted in the production of the high-Ti suite of the MLF and GPF, while further decompression into the zone of eventual breakup resulted in the concurrent generation of highly-depleted low-Ti basalts from the nascent spreading centre. This latter melting event may have lead to accumulation of residual mantle at shallow level that restricted partial melting to relatively high pressures, yielding the incompatible-element-enriched basalts of the RFF. With complete plate separation this refractory lid was removed, allowing for unimpeded upwelling. This final phase of breakup magmatism is represented by the SF, which is compositionally similar to the Tertiary lava succession on Iceland<sup>29</sup>.

We find evidence that the mantle source near the proposed axis of the Tertiary Iceland plume cooled markedly during flood volcanism and was heterogeneous, composed of depleted and less-depleted domains. This is difficult to reconcile with the northeast Atlantic being underlain by a hot, robust (>1,500 km diameter) plume head at the time of continental breakup<sup>2–4,12</sup>. If the early Tertiary plume was a narrow, rift-centred anomaly, as apparent for much of its post-breakup history<sup>9,27</sup>, then the present study would require that the convective supply of plume mantle fluctuated on a timescale greater than  $10^5$ – $10^6$  yr (ref. 6). If the plume was not rift-centred<sup>14,15</sup>, flood volcanism in east Greenland would require lateral channelling of plume mantle into the rift zone<sup>5,18</sup>. Melting in such a thin spot would be modulated by the flux of hot mantle and extension of the overlying lithosphere. This latter hypothesis could explain the close temporal and spatial association of distinct mantle domains giving rise to the high- and low-Ti lava suites of the MLF and GPF, erupted respectively from contemporaneously active continental and oceanic-type rift systems.

Active upwelling<sup>16,30,31</sup> may be an important factor in promoting the high rates of magmatic productivity along the east Greenland volcanic rifted margin despite falling mantle temperatures. Furthermore, it is possible that rifting and plume generation were not distinct events, but that rupture of the continental lithosphere induced active upwelling and an organized flow regime, producing sustained plume upwelling<sup>17</sup>. Such a plume would not originate from convective instabilities at the base of the lower or the upper mantle before breakup. But once established, active convection could involve ‘plume components’ (that is,

the composition of pooled melts for decompression melting paths beginning at various pressures of solidus intersection (that is,  $P_{\text{solidus}} = 3.5$ – $2.7$  GPa) and terminating at 0 GPa. Thin solid curves contour the final segregation pressure (that is,  $P_{\text{segregation}} = \text{top of the melting zone}$ ) and dashed curves are contours of the mean extent of melting ( $F$ ).

recycled lithosphere and  $^3\text{He}$ -rich domains)<sup>32</sup> from deep portions of the upper mantle. □

## Methods

**Sampling and analysis.** Stratigraphic control within the lava pile was assured by multi-model photogrammetry using airborne stereo-photography along the Sortebrae fjord and glacier system<sup>20</sup>, field observations, and major-element compositional data. A composite flow-to-flow profile was constructed from overlapping profiles sampled in accessible areas. Samples (~1–2 kg) were taken from the massive and least-altered parts of each flow unit, and weathered surfaces, veins and other impurities were removed before analysis. 439 samples from 331 individual lava flows were analysed on glass disks by X-ray fluorescence (XRF) for major- and some trace-elements, at the Geological Survey of Denmark and Greenland. This data set (see Supplementary Information) was screened for samples showing obvious effects of olivine accumulation (16 samples with >12 wt% MgO; all occurring in the MLF) and crustal contamination (10 samples with  $1,000(\text{Ba}/\text{Ti}) > 20$ ; mainly occurring in the MLF). The remaining 413 lavas, for which measured  $\epsilon_{\text{Nd}}$  (>+5) and  $^{87}\text{Sr}/^{86}\text{Sr}$  values (<0.7038) on a subset of samples showed no sign of significant contamination, are the focus of discussion. A subset of 150 samples representing the complete range in lava types for the Sortebrae area were analysed for REE abundances by inductively-coupled plasma mass spectrometry (ICP-MS) using the Fisons PQ2+ PlasmaQuad facility at Oregon State University on powdered material dissolved in acids. Total dissolution of samples is confirmed by a ~1:1 correlation of Zr measured by XRF and ICP-MS ( $R^2 = 0.92$ ). 87 repeat analyses of a monitor sample (GGU95358) showed that the reproducibility (total procedure) expressed as a percentage standard error relative to the average concentration of the REEs is <3.5%, that is, much less than the size of symbols used in the figures.

**Modelling mantle melting.** Forward models of polybaric melting shown in Fig. 3b assume a source mantle depleted (by 0.5% batch melting of spinel lherzolite) relative to primitive mantle<sup>21</sup>. This mantle source is used for illustration, and is consistent with positive  $\epsilon_{\text{Nd}}$  (>+5), requiring a long-term  $\text{Sm}/\text{Nd} > 0.322$ , and geochemical comparisons between east Greenland flood basalts and modern Iceland and northeast Atlantic MORB<sup>24</sup>. We note that computations using more depleted mantle sources (1–2% depleted) or primitive mantle fail to recover fundamental REE characteristics of the basalt suite, such as corresponding high  $\text{Dy}/\text{Yb}_N$  and low  $\text{La}/\text{Sm}_N$  in the MLF. The computer algorithm simulating dynamic decompression melting is from Fram

and Leshner<sup>10</sup> with the following constraints: melting proceeds by incremental non-modal batch melting over a range of pressures from the base to the top of the melting column, assuming a uniform melt productivity of 1% per kbar (refs 25, 26); REE distribution coefficients and reaction coefficients used are summarized in refs 28 and 33, and references therein. The garnet–spinel transition is assumed to be a linear change between 3 and 2.5 GPa, and the spinel–plagioclase transition is assumed to be linear between 1.1 and 0.9 GPa. The initial mineral mode for olivine:orthopyroxene:clinopyroxene:garnet is 55:15:20:10 (ref. 33). The composition of the residue is calculated after each increment (that is, one kbar) of decompression melting and adjusted for the pressure dependent phase transitions. The shape of the melting regime is defined by simple corner flow, assuming that the residual mantle accumulates along the leading edges of rifting (or drifting) continental lithosphere.

Received 25 February; accepted 26 June 1998.

- Brooks, C. K. Rifting and doming in southern East Greenland. *Nature* **244**, 23–25 (1973).
- White, R. S. & McKenzie, D. P. Magmatism at rift zones: The generation of volcanic continental margins and flood basalts. *J. Geophys. Res.* **94**, 7685–7729 (1989).
- Campbell, I. H. & Griffiths, R. W. Implications of mantle plume structure for the evolution of flood basalts. *Earth Planet. Sci. Lett.* **99**, 79–93 (1990).
- Fitton, J. G., Saunders, A. D., Norry, M. J., Hardason, B. S. & Taylor, R. N. Thermal and chemical structure of the Iceland plume. *Earth Planet. Sci. Lett.* **153**, 197–208 (1997).
- Larsen, H. C. & Saunders, A. D. Tectonism and volcanism at the SE Greenland rifted margin: a record of plume impact and later continental rapture. *Proc. ODP Sci. Res.* **152**, 502–533 (1998).
- Schilling, J.-G., Meyer, P. S. & Kingsley, R. H. Evolution of the Iceland hotspot. *Nature* **296**, 313–320 (1982).
- Elliott, T. R., Hawkesworth, C. J. & Gronvold, K. Dynamic melting of the Iceland plume. *Nature* **351**, 201–206 (1991).
- Hémond, C., Arndt, N., Lichtenstein, U., Hofmann, A. W., Oskarsson, N. & Steinthorsson, S. The heterogeneous Iceland plume: Nd–Sr–O isotopes and trace element constraints. *J. Geophys. Res.* **98**, 15833–15850 (1993).
- Wolfe, C. J., Bjarnason, I. T., VanDecar, J. C. & Solomon, S. C. Seismic structure of the Iceland mantle plume. *Nature* **385**, 245–247 (1997).
- Fram, M. S. & Leshner, C. E. Geochemical constraints on mantle melting during creation of the North Atlantic basin. *Nature* **363**, 712–715 (1993).
- White, R. S. & McKenzie, D. P. Mantle plumes and flood basalts. *J. Geophys. Res.* **100**, 17543–17585 (1995).
- Barton, A. J. & White, R. S. Crustal structure of Edoras Bank continental margin and mantle thermal anomalies beneath the North Atlantic. *J. Geophys. Res.* **102**, 3109–3129 (1997).
- Schilling, J.-G. & Noe-Nygaard, A. Faeroe–Iceland plume: rare-earth evidence. *Earth Planet. Sci. Lett.* **24**, 12–14 (1974).
- Lawver, L. A. & Müller, R. D. Iceland hotspot track. *Geology* **22**, 311–314 (1994).
- Tegner, C. *et al.* <sup>40</sup>Ar–<sup>39</sup>Ar geochronology of Tertiary mafic intrusions along the East Greenland rifted margin: Relation to flood basalts and the Iceland hotspot track. *Earth Planet. Sci. Lett.* **156**, 75–88 (1998).
- Mutter, J. C., Buck, W. R. & Zehnder, C. M. Convective partial melting 1. A model for the formation of thick basaltic sequences during the initiation of spreading. *J. Geophys. Res.* **93**, 1031–1048 (1988).
- Anderson, D. L., Zhang, Y. S. & Tanimoto, T. in *Magmatism and the Causes of Continental Break-up* (eds Storey, B. C., Alabaster, T. & Pankhurst, R. J.) 99–124 (Spec. Publ. 68, Geol. Soc. London, 1992).
- Sleep, N. H. Lateral flow of hot plume material ponded at sublithospheric depths. *J. Geophys. Res.* **101**, 28065–28083 (1996).
- Larsen, L. M., Watt, W. S. & Watt, M. Geology and petrology of the Lower Tertiary plateau basalts of the Scoresby Sund region, East Greenland. *Bull. Grønlands Geol. Unders.* **157**, 1–164 (1989).
- Pedersen, A. K., Watt, M., Watt, W. S. & Larsen, L. M. Structure and stratigraphy of the early Tertiary basalts of the Blosseville Kyst, East Greenland. *J. Geol. Soc. Lond.* **154**, 565–570 (1997).
- Sun, S. S. & McDonough, W. F. in *Magmatism in the Ocean Basins* (eds Saunders, A. D. & Norry, M. J.) 313–345 (Spec. Publ. 42, Geol. Soc., London, 1989).
- Schilling, J.-G. *et al.* Petrologic and geochemical variations along the Mid-Atlantic ridge from 29° N to 73° N. *Am. J. Sci.* **283**, 510–586 (1983).
- Klein, E. M. & Langmuir, C. H. Global correlations of ocean ridge basalt geochemistry with axial depth and crustal thickness. *J. Geophys. Res.* **92**, 8089–8115 (1987).
- Fram, M. S. & Leshner, C. E. Generation and polybaric differentiation of magmas of East Greenland Early Tertiary flood basalts. *J. Petrol.* **38**, 231–275 (1997).
- McKenzie, D. & Bickle, M. J. The volume and the composition of melt generated by extension of the lithosphere. *J. Petrol.* **29**, 625–679 (1988).
- Langmuir, C. H., Klein, E. M. & Plank, T. in *Mantle Flow and Melt Generation at Mid-ocean Ridges* (eds Morgan, J. P., Blakmun, D. K. & Sinton, J. M.) 183–280 (Am. Geophys. Union Monogr., Washington DC, 1992).
- Schilling, J.-G. Fluxes and excess temperatures of mantle plumes inferred from their interactions with migrating mid-ocean ridges. *Nature* **352**, 397–403 (1991).
- Fram, M. S., Leshner, C. E. & Volpe, A. M. Mantle melting systematics: the transition from continental to oceanic volcanism on the Southeast Greenland margin. *Proc. ODP Sci. Res.* **152**, 373–386 (1998).
- Hardason, B. S., Fitton, J. G., Ellam, R. M. & Pringle, M. S. Rift relocation—a geochemical and geochronological investigation of a paleo-rift in northwest Iceland. *Earth Planet. Sci. Lett.* **153**, 181–196 (1997).
- Hopper, J. R. *et al.* Magmatism and rift margin evolution: Evidence from northwest Australia. *Geology* **20**, 853–857 (1992).
- Holbrook, W. S. & Kelemen, P. B. Large igneous province on the US Atlantic margin and implications for magmatism during continental breakup. *Nature* **364**, 433–436 (1993).
- Hofmann, A. W. Mantle geochemistry: the message from oceanic volcanism. *Nature* **385**, 219–229 (1997).
- Leshner, C. E. & Baker, M. B. Near-solidus phase relations of mantle peridotite at 3.6–5 GPa. *Eos* **78**, 812 (1997).

**Supplementary information** is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

**Acknowledgements.** We thank our colleagues (S. Bernstein, D. K. Bird, C. K. Brooks, R. A. Duncan, J. Helgason, P. B. Kelemen, P. S. Neuhoff, P. O'Day and M. Storey) involved in sampling of East Greenland flood basalts in 1995, coordinated by A. K. Pedersen and T. F. D. Nielsen. Discussions with H. C. Larsen, T. F. D. Nielsen, C. K. Brooks, M. Storey, and J. R. Hopper helped to mature the ideas presented here. The Danish National Research Foundation and the US NSF supported this work.

Correspondence and requests for materials should be addressed to C.E.L. at the University of California (e-mail: [lesher@geology.ucdavis.edu](mailto:lesher@geology.ucdavis.edu)).

## Apostatic selection by blue jays produces balanced polymorphism in virtual prey

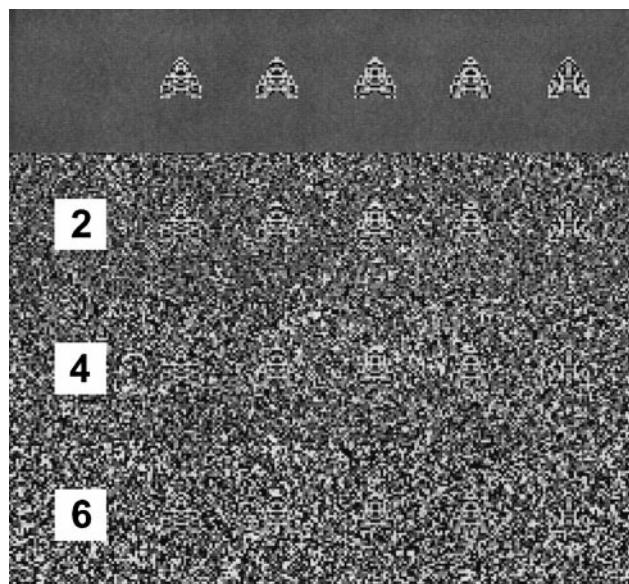
Alan B. Bond\*† & Alan C. Kamil\*†‡

\* School of Biological Sciences, † Nebraska Behavioral Biology Group, and

‡ Department of Psychology, University of Nebraska, Lincoln, Nebraska 68588-0118, USA

Apostatic selection, in which predators overlook rare prey types while consuming an excess of abundant ones, has been assumed to contribute to the maintenance of prey polymorphisms<sup>1–3</sup>. Such an effect requires predators to respond to changes in the relative abundance of prey, switching to alternatives when a focal prey type becomes less common<sup>4,5</sup>. Apostatic selection has often been investigated using fixed relative proportions of prey<sup>1,6</sup>, but its effects on predator–prey dynamics have been difficult to demonstrate<sup>7</sup>. Here we report results from a new technique that incorporates computer-generated displays<sup>8,9</sup> into an established experimental system, that of blue jays (*Cyanocitta cristata*) hunting for cryptic *Catocala* moths<sup>10</sup>. Digital prey images from a virtual population are presented to predators. The relative numbers that escape detection determine the subsequent abundance of each prey type. If apostatic selection does promote stability, the system should converge on an equilibrium in which each prey type appears at a characteristic abundance. Our results show that the detection of cryptic prey does involve apostatic selection, and that such selection can function to maintain prey polymorphism.

We began by scanning a photograph<sup>11</sup> of the dark morph of *Catocala relictata*, a moth commonly preyed on by jays. Four novel morphs, moths 1–4, were generated by random modification of this prototype, whereas moth 5 was obtained by similarly processing a photograph of *C. relictata*. The resulting digital moths were bilaterally symmetrical, roughly triangular, and disparate in appearance, at



**Figure 1** The five types of digital moths used in these experiments (moth 1 to moth 5, from left to right), presented against backgrounds of three levels of crypticity to illustrate the difficulty of the detection task. When projected on a computer monitor, the images of the moths were ~6 mm high. The image can be viewed in more detail at <http://niko.unl.edu/~kamil/moths/vpmoths.htm>.



least to human eyes. They were rendered difficult to detect by presenting them on backgrounds that approximated their pixel intensities and pattern granularity<sup>12</sup> (Fig. 1).

Six hand-reared blue jays with experience in similar tasks searched for prey on a computer screen in a series of discrete trials. Trials were either prey trials, with a moth overlaid on the background at a random location, or no-prey trials, with no moth present. If the bird found and pecked at a moth, it received a food reward. If not, it could proceed rapidly to the next trial by pecking a central green disk. Each bird received 36 prey trials and 84 no-prey trials per day in random order. The 216 moths exposed to potential predation each day were drawn at random without replacement from a virtual population of 240 moths. Detected moths were considered 'killed' and were removed from the pool. At the beginning of the next day, the population was regenerated to 240, maintaining the relative abundance shown by the surviving prey. Thus, each day formed a 'generation', and the effects of predation were reflected in the frequency of the different morphs the following day. The only limits on free variation were the initial relative proportions of the prey morphs, which we set at the beginning of each replication, and the constraint that no morph was allowed a frequency of less than five individuals at the start of a day.

The experiment began with a founding population consisting of equal numbers of moths 1–3. Over the subsequent 50 days, the abundance of moth 1, the most cryptic morph, increased from its initial value to a maximum of ~180 individuals before levelling off (Fig. 2a). The numbers of moths 2 and 3, in contrast, dropped to around 30 individuals each. We then began a second replication by raising moth 2 to 180 of the 240 individuals in the population and dropping both moths 1 and 3 to 30 individuals. After 50 days, equilibrium levels similar to those in the first replication were attained (Fig. 2a). We then conducted a third replication by maximizing the initial relative abundance of moth 3. In all three replications, equilibrium population levels appeared to be attained with 30 generations, with ~75% of the population consisting of moth 1 and 12.5% each of moths 2 and 3.

In these three replications, no morph ever declined to fewer than five individuals, and we never had to intervene to prevent extinction. Both population levels and detectability (Fig. 2b) exhibited a slow oscillation with a period, shown by autocorrelation analysis, of about ten generations ( $r = 0.11$ ,  $P < 0.04$ ). For this ten-generation cycle, the mean phase difference between population numbers and

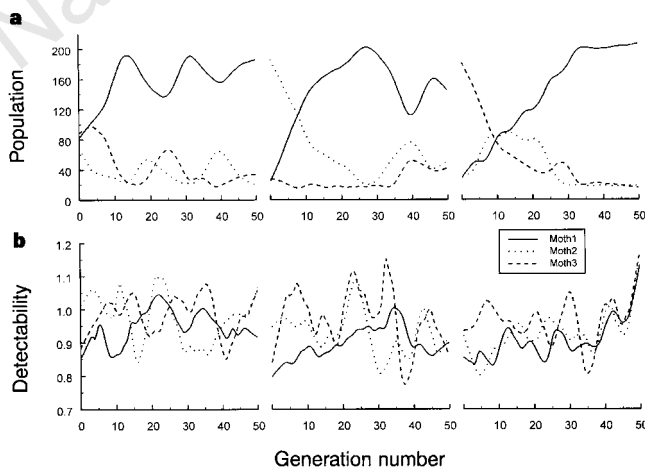
detectability was  $+0.3\pi$ , which indicates that peak detectability lagged behind peak population by ~1.5 generations. These results, together with the stable equilibrium that follows perturbation, indicate that negative feedback between population numbers and detectability is sustaining the polymorphism.

The functional response of a predator is the relationship between the proportion of each prey morph in the diet and the proportion available in the environment<sup>13</sup>. A sigmoid functional response, in which the proportion in the diet is amplified at high relative abundances and diminished at low ones, is a defining criterion for apostatic selection<sup>2</sup>. In Fig. 3 we display the residuals from the expected linear function, that is, plots of the difference between proportion of each prey morph taken and proportion presented as a function of the proportion presented. Apostatic selection predicts that the residuals should increase with the proportion presented, corresponding to overselection at high relative abundances.

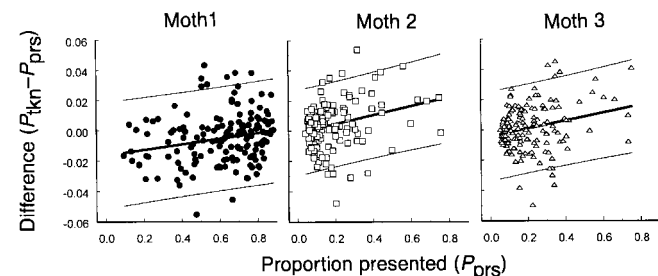
Linear regression analysis yielded significant positive slopes for all three morphs, confirming the prediction. Analysis of covariance found no significant differences among slopes. The intercept was significantly less than 0.0 only for moth 1 ( $P < 0.01$ ), indicating that this more cryptic morph was often overlooked when it was rare. Although it is possible that these effects resulted from a simple preference for the most abundant prey, previous work has shown repeatedly that a predator's ability to detect cryptic prey varies with the encounter frequency, a phenomenon often termed 'searching image'<sup>10,14–17</sup>. In a separate study, jays searching for cryptic digital moths showed clear searching-image effects when presentation probabilities were manipulated systematically (A.B.B. & A.C.K., in preparation).

At equilibrium, frequency-dependent selection should result in equal fitnesses for all morphs<sup>18</sup>, expressed here as equal risks of detection. This provides a quantitative test of convergence to equilibrium. We eliminated oscillatory noise by extracting the linear trend of detectability as a function of generation for each morph in each replication. The predicted values for generation 0 from the trend analysis showed a far larger variance than those for generation 50 ( $F(8,8) = 4.10$ ,  $P = 0.03$ ). Detectability values for the first three morphs thus converged over the course of each replication, confirming the approach to equilibrium.

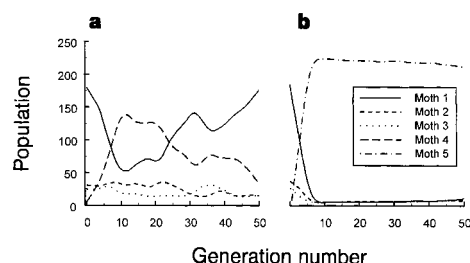
Frequency-dependent selection should also maximize the average fitness of the prey population at equilibrium<sup>18</sup>. We examined changes in foraging success over the course of a replication by



**Figure 2** Population numbers and prey detectability of three prey morphs in three successive replications of the virtual prey procedure. **a**, Population numbers of the three morphs. Curves were smoothed with weighted least squares, using an eight-generation window. **b**, Prey detectability, expressed as the arcsine of the proportion of correct responses to prey trials, in the same three replications. Curves were smoothed as in **a**.



**Figure 3** Functional response curves for the three morphs in the first three replications. Each data point indicates the relationship for a single daily session. The abscissa is  $P_{\text{prs}}$ , the proportion of prey trials that included the given moth type; the ordinate is the difference between  $P_{\text{tkn}}$ , the proportion of that moth detected ('taken') by the birds, and the proportion presented. The regression line indicating the association between the difference and  $P_{\text{prs}}$  is shown, together with its 95% confidence limits. The slopes for each prey type were: moth 1,  $\beta = 0.018$ ,  $P = 0.009$ ; moth 2,  $\beta = 0.030$ ,  $P = 0.0004$ ; moth 3,  $\beta = 0.026$ ,  $P = 0.002$ .



**Figure 4** Population numbers of four morphs in the last two replications of the virtual predation procedure. Curves were smoothed with weighted least squares, using an eight-generation window. Replication 4, shown in **a**, included moths 1–3 and moth 4; replication 5, shown in **b**, included moths 1–3 and moth 5.

dividing number of prey detected within each generation by total search time, summed across birds and trials. The linear trend of foraging success as a function of generation number yielded significant negative slopes for all replications (replication 1:  $\beta = -0.0014$ ,  $P = 0.02$ ; replication 2:  $\beta = -0.0019$ ,  $P = 0.003$ ; replication 3:  $\beta = -0.0014$ ,  $P = 0.014$ ), reflecting the higher relative abundance of the most cryptic morph at equilibrium. By tending to search for the morph that is currently most abundant, predators improve their ability to detect that morph, maximizing their short-term foraging success. However, the long-term consequence is to generate a distribution of prey abundances that yields a significantly reduced rate of return.

The strongest effects of apostatic selection should be evident with novel prey. Tinbergen<sup>14</sup> found that insectivorous birds generally overlooked newly emergent prey until they had attained high population levels. At some point, however, the birds would start concentrating on the new prey, driving down abundance. We simulated this situation in the fourth and fifth replications. In replication 4, moths 1–3 were maintained at their proportional equilibrium abundances, and seven individuals of moth 4 were added to the population. Moth 4 was initially detected infrequently, with an average detectability in the first three generations of only 0.64. By the tenth generation, however, detectability reached levels comparable to the other three morphs, and the population of moth 4 began to decline, reaching levels similar to that of moths 2 and 3 (Fig. 4a).

In replication 5, we followed the same procedure, introducing seven individuals of moth 5, which was exceedingly difficult to detect. Moth 5 was completely overlooked until generation 7, at which point it constituted 94% of the population (Fig. 4b). One jay began feeding on moth 5 at this time, and a second started at generation 14. Both birds rapidly increased their detectability on this morph to 0.9–1.0. However, no other birds learned to detect moth 5, even after 50 generations and, consequently, moth 5 sustained maximal population levels through the replication (Fig. 4b). The other morphs would have been driven to extinction if they had not been repeatedly brought up to the minimum of five individuals at the start of each generation.

These last results show that stable equilibrium is not an invariable result of visual search for cryptic, polymorphic prey. If the prey are too conspicuous, frequency-dependent effects will be small, and random drift will eventually result in local extinctions<sup>1,6,15</sup>. Alternatively, novel prey that are too cryptic or predators that are too inflexible in searching strategies can apparently also result in fixation of more cryptic morphs<sup>3</sup>. To our knowledge, the dynamics of balanced polymorphism, although long established in theory, have never before been demonstrated empirically in a predator–prey system. Because the virtual prey technique incorporated the reciprocal interaction between the predator's searching behaviour and the relative abundance of different prey morphs, we have been

able to show that changes in prey detection alone can produce stability and can maintain prey polymorphism. □

## Methods

**Stimuli.** The initial digital image was reformed into the resting pose that was typical of the species, with the head orientated upwards, exposing only the cryptic upper surfaces of the forewings. It was then reduced to fit in a sixteen-pixel square, histogram-normalized, and converted to grey-scale at SVGA resolution ( $640 \times 480 \times 64$  levels of grey). Corresponding pixels were averaged about the central axis to ensure bilateral symmetry. Novel morphs were generated from the initial prototype by inverting the brightness values of a randomly selected subset of 16% of the wing pixels. Cryptic backgrounds were generated based on the pooled frequency distribution of brightness values from the moth images. This distribution was roughly bimodal: the highest concealing effect, for human evaluators, was obtained with equal-variance peaks with modes of six and 56. To vary task difficulty, we created backgrounds that were random mixtures of pixels from this bimodal distribution and another, unimodal distribution centred around 30, against which the digital moths seemed far more conspicuous. The difficulty level was coded as an integer, with 1 indicating that 10% of the pixels came from the bimodal distribution, 2 meaning 20% bimodal, and so on. In the final step, a fractal texture was imposed on the background, increasing the granularity to correspond more closely to the patterns exhibited by the digital moths.

**Subjects and procedures.** The blue jays, captured as nestlings and hand-reared, had all participated in visual detection experiments, including studies using the same backgrounds and similar prey stimuli. Each trial began by displaying two  $9.5 \times 13$  cm fields of cryptic background with a 2.7-cm green disk between them. Equal numbers of each of five levels of background difficulty (4–8 on the scale shown in Fig. 1) were used each day. The screen was surrounded with an infrared touch frame; a single peck to the screen ended the trial and caused the display to go blank. During a prey trial, a peck directed at the moth was rewarded with a piece of mealworm, and a new trial followed after 12 s. During either prey or no-prey trials, pecks to the green disk reduced the time to the next trial to 6 s, whereas pecks to the background away from any moth image increased it to 30 s. Erroneous responses to the background, which were uncommon, were thus discouraged by long delays before beginning the next trial.

Received 27 May; accepted 29 July 1998.

- Allen, J. A. Frequency-dependent selection by predators. *Phil. Trans. R. Soc. Lond. B* **319**, 485–503 (1988).
- Clarke, B. C. in *Taxonomy and Geography* (ed. Nichols, D.) 47–70 (Systematics Association, Oxford, 1962).
- Clarke, B. C. The evidence for apostatic selection. *Heredity* **24**, 347–352 (1969).
- Murdoch, W. W. Switching in general predators: experiments on predator specificity and stability of prey populations. *Ecol. Monogr.* **39**, 335–354 (1969).
- Murdoch, W. W. & Oaten, A. Predation and population stability. *Adv. Ecol. Res.* **9**, 1–131 (1974).
- Cooper, J. M. Apostatic selection on prey that match the background. *Biol. J. Linn. Soc.* **23**, 221–228 (1984).
- Cooper, J. M. & Allen, J. A. Selection by wild birds on artificial dimorphic prey on varied backgrounds. *Biol. J. Linn. Soc.* **51**, 433–446 (1994).
- Glanville, P. W. & Allen, J. A. Protective polymorphism in populations of computer-simulated moth-like prey. *Oikos* **80**, 565–571 (1997).
- Plaisted, K. C. & Mackintosh, N. J. Visual search for cryptic stimuli in pigeons: implications for the search image and search rate hypotheses. *Anim. Behav.* **50**, 1219–1232 (1995).
- Pietrewicz, A. T. & Kamil, A. C. Search image formation in the blue jay (*Cyanocitta cristata*). *Science* **204**, 1332–1333 (1979).
- Sargent, T. D. *Legion of Night: The Underwing Moths* (Univ. Massachusetts Press, Amherst, 1976).
- Ender, J. A. Progressive background matching in moths, and a quantitative measure of crypsis. *Biol. J. Linn. Soc.* **22**, 187–231 (1984).
- Holling, C. S. The functional response of predators to prey density and its role in mimicry and population regulation. *Mem. Entomol. Soc. Can.* **45**, 1–60 (1965).
- Tinbergen, N. The natural control of insects of pinewoods I. Factors influencing the intensity of predation by songbirds. *Arch. Néerland. Zool.* **13**, 265–343 (1960).
- Bond, A. B. Visual search and selection of natural stimuli in the pigeon: the attention threshold hypothesis. *J. Exp. Psychol. Anim. Behav. Process.* **9**, 292–306 (1983).
- Blough, P. M. Selective attention and search images in pigeons. *J. Exp. Psychol. Anim. Behav. Process.* **17**, 292–298 (1991).
- Langley, C. M. Search images: selective attention to specific visual features. *J. Exp. Psychol. Anim. Behav. Process.* **22**, 152–163 (1996).
- Wallace, B. *Basic Population Genetics* 229–230 (Columbia Univ. Press, New York, 1981).

**Acknowledgements.** We thank K. Cheng, J. Ender, A. Joern, S. Shettleworth, A. Zera, W. Wagner, J. Allen and J. Diamond for comments and suggestions. This research was supported by a grant from The National Science Foundation.

Correspondence and requests for materials should be addressed to either A.B.B. (e-mail: abond@unl.edu) or A.C.K. (kamil@niko.unl.edu).

# Imaging unconscious semantic priming

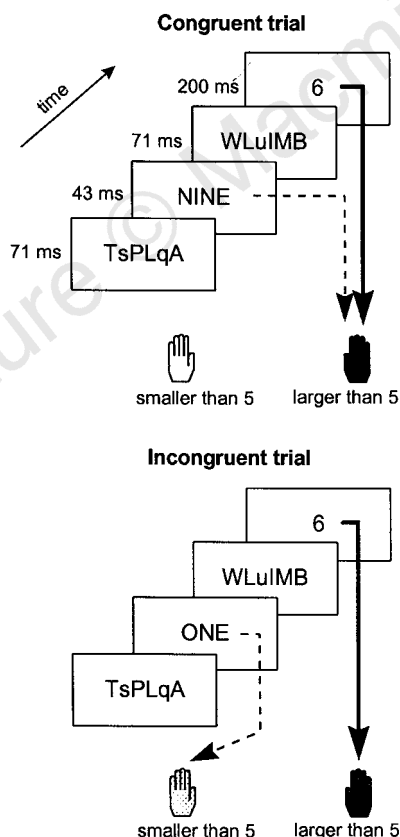
Stanislas Dehaene\*, Lionel Naccache\*, Gurvan Le Clec'H\*, Etienne Koechlin\*, Michael Mueller\*, Ghislaine Dehaene-Lambertz†, Pierre-François van de Moortele‡ & Denis Le Bihan‡

\* INSERM U.334, Service Hospitalier Frédéric Joliot, CEA/DRM/DSV, 4 Place du Général Leclerc, 91401 Orsay, France

† Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS/CNRS, 75270 Paris cedex 06, France

‡ Service Hospitalier Frédéric Joliot, CEA/DRM/DSV, 91401 Orsay, France

Visual words that are masked and presented so briefly that they cannot be seen may nevertheless facilitate the subsequent processing of related words, a phenomenon called masked priming<sup>1,2</sup>. It has been debated whether masked primes can activate cognitive processes without gaining access to consciousness<sup>3–5</sup>. Here we use a combination of behavioural and brain-imaging techniques to estimate the depth of processing of masked numerical primes. Our results indicate that masked stimuli have a measurable influence on electrical and haemodynamic measures of brain activity. When subjects engage in an overt semantic comparison task with a clearly visible target numeral, measures of covert motor activity indicate that they also unconsciously apply the task instructions to an unseen masked numeral. A stream of perceptual, semantic and motor processes can therefore occur without awareness.



**Figure 1** Experimental design. In each trial, the following four visual stimuli appeared successively, centred on the same screen location: a random-letter-string mask, a prime number, another mask, and a target number. Subjects were not informed of the presence of the prime. They were simply told that a 'signal' preceded the target number that they had to classify as larger or smaller than 5. Half of the trials were of the congruent type (prime and target both falling on the same side of 5), and half were incongruent.

We presented a numeral between 1 and 9, the prime, to subjects for a very short duration (43 ms). The prime was masked by two nonsense letter strings, and followed by another numeral, the target (Fig. 1). Under these conditions, even when subjects focused their attention on the prime, they could neither reliably report its presence or absence nor discriminate it from a nonsense string (Table 1). Nevertheless, we show here that the prime is processed to a high cognitive level.

We asked subjects to perform a simple semantic categorization task on the target numeral. Subjects were asked to press a response key with one hand if the target was larger than 5, and with the other hand if the target was smaller than 5. Unknown to them, each target number was preceded by a masked prime which was varied from trial to trial so it too could be larger or smaller than 5. In some trials the prime was congruent with the target (both numbers fell on the same side of 5), and in other trials it was incongruent (one number being larger than 5, and the other being smaller; Fig. 1). We first established that prime–target congruity has a significant influence on behavioural, electrical and haemodynamic measures of brain function. We then showed that the interference between prime and target can be attributed to a covert, prime-induced activation of motor cortex, a response bias that must be overcome in incongruent trials. This indicates that the prime was unconsciously processed according to task instructions, all the way down to the motor system.

The effect of prime–target congruity on behaviour is shown in Fig. 2. Subjects responded more slowly in incongruent trials than in congruent trials ( $P < 0.0001$ ). All 12 subjects showed a positive priming effect, ranging from 2 to 43 ms (average 24 ms, s.d. 13.5 ms). Furthermore, the entire response time distribution was shifted by  $\sim 24$  ms in incongruent trials compared with congruent trials. Thus, there was no evidence that the effect was found in only a small proportion of subjects or a small number of trials with a distinct distribution of response times.

Two characteristics of the priming effect link it to a semantic level of analysis. First, we varied the notations used for the prime number and for the target number independently. Either number could be presented as an arabic digit (for example, 1 or 4) or as a written word (for example, ONE or FOUR). Although the target notation had a significant main effect on response times ( $P < 0.0001$ ), the amount of priming itself was similar under all conditions of notation and did not interact with target notation, prime notation, or notation change. Priming remained significant even when the notations of the prime and target numbers differed (for example, prime 4, target ONE;  $P = 0.0006$ ). Thus, priming occurred at a notation-independent level of numerical representation<sup>6</sup>. Second, priming remained significant even after trials with repeated numbers (for example,

**Table 1** Experimental measures of prime awareness

|                  | Prime duration (ms) |      |      |        |        |        |
|------------------|---------------------|------|------|--------|--------|--------|
|                  | 0                   | 29   | 43   | 57     | 114    | 200    |
| <b>Task 1</b>    |                     |      |      |        |        |        |
| Hit rate (%)     | 4.2                 | 10.4 | 12.5 | 25.0   | 85.4   | 97.9   |
| False alarms (%) | 4.2                 | 7.3  | 7.3  | 3.1    | 5.2    | 1.0    |
| $d'$             | 0.0                 | 0.20 | 0.30 | 1.19** | 2.68** | 4.36** |
| <b>Task 2</b>    |                     |      |      |        |        |        |
| Hit rate (%)     | 28.6                | 40.2 | 49.1 | 46.4   | 78.6   | 95.5   |
| False alarms (%) | 34.8                | 32.1 | 41.1 | 30.4   | 28.6   | 16.1   |
| $d'$             | -0.17               | 0.22 | 0.20 | 0.42*  | 1.36** | 2.69** |

In two control experiments, subjects were fully informed of the precise structure of the stimuli and were then presented with trials with numerical primes intermixed with trials in which the primes were omitted (explicit detection, task 1; six subjects, 96 trials per cell) or replaced by random strings with the same number of characters (number versus letter-string discrimination, task 2; seven subjects, 112 trials per cell). Prime duration was systematically varied. At the prime duration used in the main experiments (43 ms), subjects consistently reported not seeing the numerical primes (task 1), did not respond differently to prime-absent and prime-present trials (task 1) and were unable to discriminate numerical primes from letter strings (task 2). Discrimination performance, as measured by  $d'$ , a bias-free measure of stimulus discriminability derived from signal-detection theory, began to deviate from chance only for a prime duration of 57 ms or more ( $\chi^2$  test; asterisk indicates  $P < 0.05$ ; double asterisk indicates  $P < 0.001$ ).

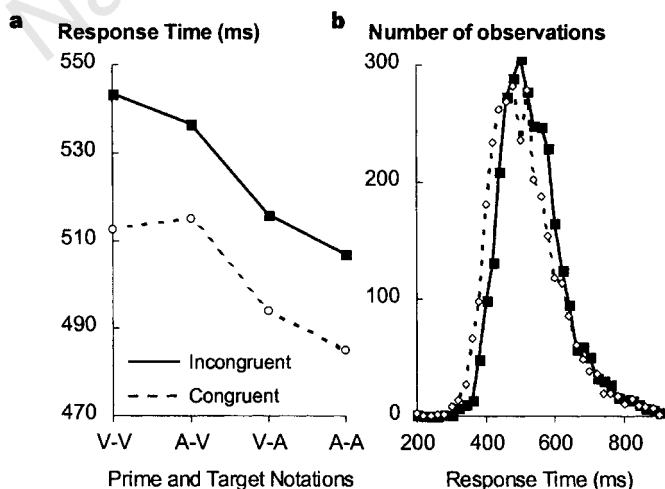


prime ONE, target 1) were excluded from the analysis ( $P = 0.001$ ). Hence, priming did not simply reflect a word repetition effect, but depended on the similarity of the prime and target numbers at a semantic level (larger or smaller than 5).

Event-related potentials (ERPs) recorded during the task also showed a prime–target congruity effect. The central positivity, which culminated at about the time of the subject's response and is thought to index post-perceptual processing<sup>7</sup>, was delayed by ~24 ms in incongruent trials compared with congruent trials (Fig. 3a). We proposed that the slower responses in incongruent trials might be due to response competition. Subjects would unconsciously apply the task instructions to the prime, would therefore categorize it as smaller or larger than 5, and would even prepare a motor response appropriate to the prime (Fig. 1). In incongruent trials, this prime-induced covert motor activation would mismatch with the overt response required to the target, resulting in response competition and hence slower response times relative to congruent trials.

According to this theory, we should be able to detect an early covert motor activation on the correct response side during congruent trials, and on the incorrect response side during incongruent trials. We tested this prediction using the lateralized readiness potential (LRP), an ERP measure that indexes the activation of lateralized motor circuits<sup>8</sup> (Fig. 3b). The LRP can detect low levels of covert response activation that do not necessarily result in an overt motor response<sup>8–10</sup>. As it unfolds in time, the LRP departs from zero as soon as task-relevant information becomes available to bias the motor response towards the left or right hand. Conventionally, positive deflections indicate response preparation on the correct side, whereas negative deflections indicate a transient covert activation on the incorrect response side. Here, the LRP revealed the predicted covert prime-induced motor preparation. In response-locked averages, before the large positive deviation which reflected the activation of motor circuits on the correct response side, the LRP showed a significant difference in the predicted direction between congruent and incongruent trials (one-tailed  $P = 0.015$ ). In incongruent trials, the LRP showed a significant negative deviation (one-tailed  $P = 0.005$ ), indicating motor preparation on the incorrect side of response, whereas in congruent trials a nonsignificant positive deviation occurred (one-tailed  $P = 0.22$ ). Hence, a period of covert prime-induced response competition preceded the overt execution of the correct response.

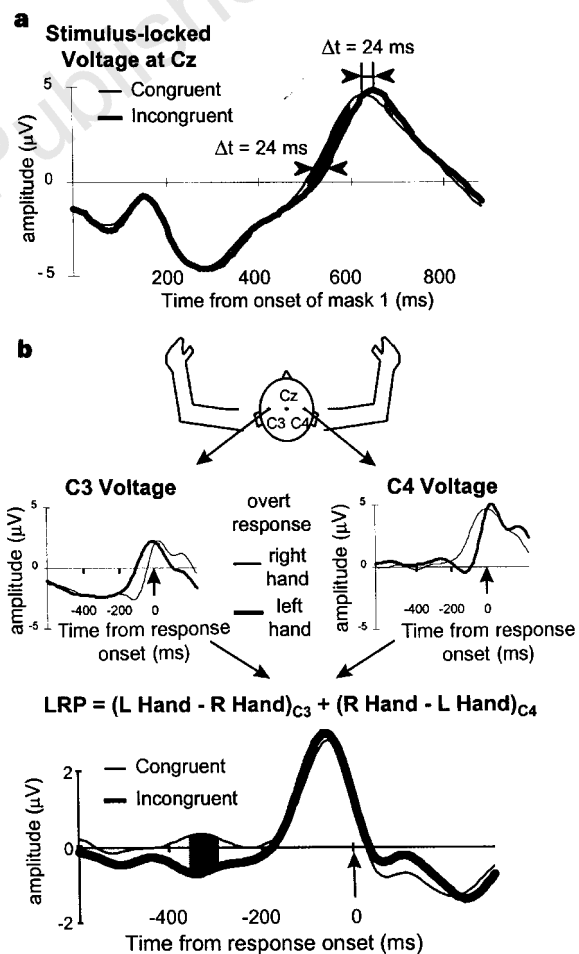
Covert prime processing could be observed even more directly.



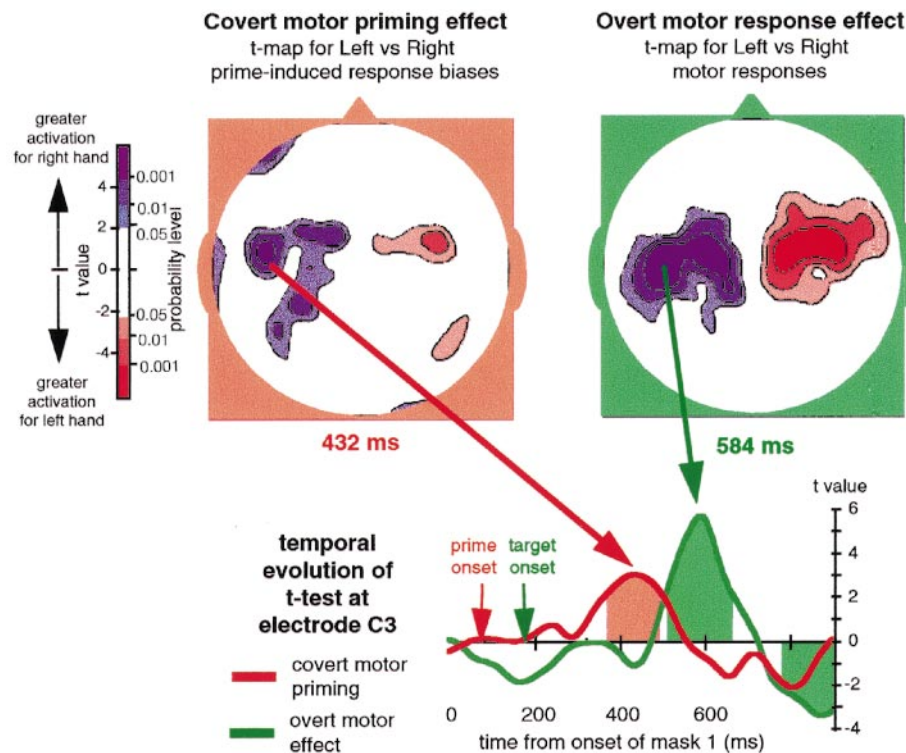
**Figure 2** Behavioural priming effect. **a**, Average correct response times recorded during the ERP experiment are plotted as a function of prime–target congruity for different prime and target notations (A, Arabic; V, verbal). **b**, The distribution of correct responses showed a rightward shift in incongruent trials relative to congruent trials (bin size 20 ms).

Because the primes and targets were varied independently in the experimental list, we could test their impact on ERP recordings separately. Primes that induced a covert left-hand or right-hand bias produced a distinct pattern of brain activity over the left and right motor cortices (Fig. 4). Primes that were associated with the left hand during a given block caused a contralateral right-hemispheric negativity, and primes that were associated with the right hand caused a left-hemispheric negativity. This covert motor priming effect had a similar scalp topography to the overt motor effect that was found before the actual motor response to the target, but it was smaller and arrived earlier.

Because ERPs have a notoriously imprecise spatial resolution, we wanted to confirm that covert priming originated at least in part from motor circuits using the spatially accurate method of functional magnetic resonance imaging (fMRI). Response times recorded during fMRI recording replicated the prime–target congruity effect ( $P = 0.0027$ ; effect size 20 ms). In fMRI, however, trials were now separated by 14 s, during which the rise and fall of the haemodynamic signal was measured in the whole brain every 2 s.



**Figure 3** ERP measures of prime–target congruity. **a**, A late positivity (P3) recorded from the vertex from electrode Cz showed a significant delay in incongruent trials relative to congruent trials (shaded areas,  $P < 0.05$ ). **b**, Derivation of the lateralized readiness potential (LRP). Individual trials were averaged in synchrony with the time of the key press, thus suppressing the effects of response delays. Electrodes C3 and C4 showed large voltage differences in opposite directions (top two graphs) before left-hand versus right-hand responses, reflecting the activation of the underlying motor circuits. The LRP (bottom) is the average of the differences at C3 and at C4, calculated according to the formula shown. In incongruent trials, before the main positive-going waveform reflecting overt response preparation, the LRP was significantly more negative than in congruent trials (shaded area,  $P < 0.05$ ), reflecting covert motor priming.



**Figure 4** ERP measures of covert and overt motor activation. At each electrode site, two independent *t*-tests were performed on scalp voltages. The first test compared trials with overt target-induced right-hand or left-hand responses (green curve and top right map). The second test compared trials with primes inducing a covert left-hand or right-hand bias (orange curve and top left map). Statistical parameter maps in polar coordinates (top) show colour-coded *t*-test values at each site on the scalp, at the delay at which the effect was maximal. The

sign and topography of the covert motor priming effect and of the overt motor response effect are similar, indicating that primes and targets were processed in a similar way according to task instructions. The bottom curve shows the temporal evolution of the two independent *t*-tests at electrode site C3, positioned over the left motor cortex (coloured areas,  $P < 0.05$ ). The covert effect preceded the overt effect by 162 ms, a delay roughly comparable to the interval between the onsets of the prime and target (114 ms).

Although this coarse haemodynamic measure cannot resolve the small activation delays associated with masked priming, we reasoned that it should be proportional to the total brain activity accumulated during a given trial, and should therefore reflect the sum of overt and covert activation in motor areas. We therefore extracted the fMRI signal profile from the left and right motor cortices, and used it to derive an index of lateralized motor activation analogous to the LRP, the lateralized bold response (LBR; Fig. 5). This measure showed a highly significant positive peak following each motor response. The LBR was smaller in incongruent trials than in congruent trials. The direction of this effect is identical to that seen in the LRP (Fig. 3b). Both effects indicate a significant prime-induced activation on the wrong response side on incongruent trials relative to congruent trials, diminishing the overall size of the activation on the correct motor side. Electrical and haemodynamic measures of covert masked priming were complementary: fMRI localized the priming effect to motor cortex, but was insensitive to its time course, whereas ERPs pinpointed the priming effect to a small window of time before the overt target-related motor activation.

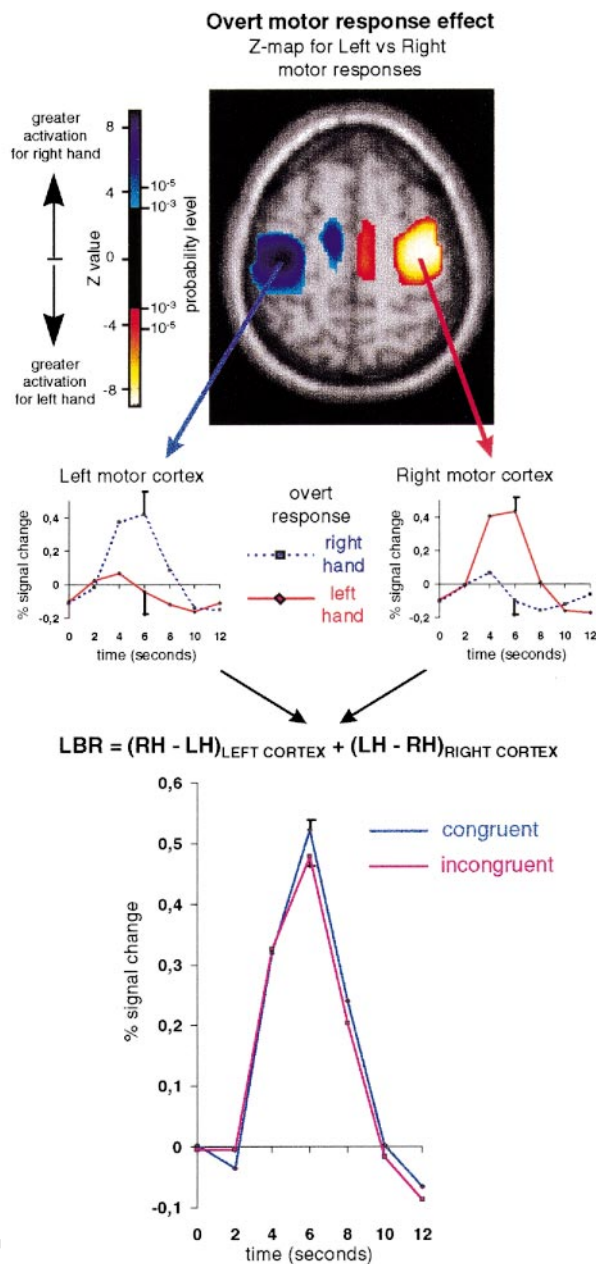
Our results resolve the issue of the depth of processing of masked primes<sup>3</sup>. First, the results show that the processing of masked primes is accompanied by measurable modifications of electrical brain activity and of cerebral blood flow. This concurs with the observation of a modulation of amygdala activity by masked visual faces<sup>11,12</sup>. As shown previously<sup>13,14</sup>, brain imaging now has the potential to image unconscious cerebral processing. Second, unconscious activity is not confined to brain areas involved in sensory processing. Even areas involved in motor programming were covertly activated here, depending on the side of the motor response that subjects should have made if they had responded to the primes according to the task instructions. Because this motor parameter was determined by whether the prime was

larger or smaller than 5, the prime must have been categorized at the semantic level. By showing that a large amount of cerebral processing, including perception, semantic categorization and task execution, can be performed in the absence of consciousness, our results narrow down the search for its cerebral substrates. □

## Methods

**Procedure.** All experiments were approved by the French ethical committee for biomedical research, and subjects gave informed consent. The stimulus set consisted of 64 pairs of prime and target numbers 1, 4, 6 and 9, each in either Arabic or spelled-out format. Subjects performed the number-comparison task twice in counterbalanced order. In one block, the instruction was to press the right-hand key for targets larger than 5 and the left-hand key for targets smaller than 5. In another block, the opposite instruction was used. Within each block, subjects received initial training (ERPs, 16 trials; fMRI, 25 trials) before the experimental session (ERPs, 256 trials; fMRI, 64 trials).

**Event-related potentials.** Twelve subjects were tested (six males; mean age 25; one other subject was rejected because of excessive motion). We presented a total of 512 stimuli at a 3-s rate on a standard PC-compatible SVGA screen (EGA mode, 70 Hz refresh rate). The electroencephalogram was digitized at 125 Hz from 128 scalp electrodes referenced to the vertex<sup>15</sup>, for a 2,048-ms period starting 400 ms before the onset of the first mask. We rejected trials with incorrect responses, voltages exceeding  $\pm 100 \mu\text{V}$ , transients exceeding  $\pm 50 \mu\text{V}$ , electro-oculogram activity exceeding  $\pm 70 \mu\text{V}$ , or response times outside a 250–1,000-ms interval. The remaining trials were averaged in synchrony either with stimulus or with response onset, digitally transformed to an average reference, band-pass filtered (0.5–20 Hz), and corrected for baseline over a 400-ms window before stimulus onset (similar results were observed with the raw, unfiltered data). Experimental conditions were compared by sample-by-sample *t*-tests on electrodes C3, C4 and Cz, with a criterion of  $P < 0.05$  for five consecutive samples. Two-dimensional maps of scalp voltage and *t*-values were constructed by spherical spline interpolation<sup>16</sup>.



**Figure 5** fMRI measure of motor priming. Top, voxels that showed significant differences in bold signal intensity between overt left-hand and right-hand responses are coded using the colour scale at left. The two voxels with the most significant overt motor effects were located in the left and right precentral cortex (Talairach coordinates  $-39, -21, 66$  and  $39, -15, 63$ ). Centre, plots of the average fMRI signal of the two voxels as a function of time show activation for contralateral movements following a haemodynamic delay of about 4–8 s (planned contrast for an increase in left–right differences from baseline (time points 0 and 2 s) to activated state (time points 4, 6 and 8 s): left motor cortex,  $F(1, 8) = 20.1$ ,  $P = 0.0021$ ; right motor cortex,  $F(1, 8) = 27.0$ ,  $P = 0.0001$ ). Bottom, it was therefore possible to construct an fMRI measure similar to the event-related LRP, indexing the time course of overt motor activation, which we term the lateralized bold response (LBR). The LBR was significantly larger (+9%) in congruent trials than in incongruent trials (interaction of the congruity factor with the baseline-to-activated-state contrast,  $F(1, 8) = 6.23$ ,  $P = 0.037$ ).

**Functional magnetic resonance imaging.** Nine new subjects were tested (seven males; mean age 26). We used an event-related design<sup>17</sup>. We presented a list of 128 randomly intermixed stimuli through mirror glasses and an active matrix video projector (EGA mode, 70 Hz refresh rate), with a 14-s inter-stimulus interval. In each trial, stimulus onset was synchronized with the

acquisition of the first slice in a series of seven volumes of eighteen slices each. We used a gradient-echo echo-planar imaging sequence sensitive to brain oxygen-level-dependent contrast (18 contiguous axial slices, 6-mm thickness, repetition time/echo time = 2,000/40 ms, in-plane resolution  $3 \times 4 \text{ mm}^2$ ,  $64 \times 64$  matrix) on a 3-Tesla whole-body system (Bruker). High-resolution anatomical images (three-dimensional gradient-echo inversion-recovery sequence, inversion time = 700 ms, repetition time = 1,600 ms, field of view =  $192 \times 256 \text{ mm}^2$ , matrix =  $256 \times 128 \times 256$ , slice thickness = 1 mm) were also acquired.

Analysis was done with SPM96 software. Images were corrected for subject motion, normalized to Talairach coordinates using a linear transform calculated on the anatomical images, smoothed (full-width at half-maximum = 15 mm), and averaged to define four types of event (congruent or incongruent  $\times$  left-hand or right-hand response). Images from all nine subjects were then analysed together. We used the generalized linear model to model the intensity level of each pixel as a linear combination, for each subject and each event type, of two activation functions with haemodynamic lags 4 and 7 s, thus allowing for differences in acquisition and activation times across slices and brain regions. We used a voxelwise significance level of 0.001, corrected to  $P < 0.05$  for multiple comparisons across the brain volume.

Received 21 May; accepted 17 July 1998.

1. Marcel, A. J. Conscious and unconscious perception: experiments on visual masking and word recognition. *Cogn. Psychol.* **15**, 197–237 (1983).
2. Forster, K. I. & Davis, C. Repetition priming and frequency attenuation in lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* **10**, 680–698 (1984).
3. Holender, D. Semantic activation without conscious identification in dichotic listening, parafoveal vision and visual masking: a survey and appraisal. *Behav. Brain Sci.* **9**, 1–23 (1986).
4. Cheesman, J. & Merikle, P. M. Priming with and without awareness. *Percept. Psychophys.* **36**, 387–395 (1984).
5. Merikle, P. M. Perception without awareness: critical issues. *Am. Psychol.* **47**, 792–796 (1992).
6. Dehaene, S. & Akhavein, R. Attention, automaticity and levels of representation in number processing. *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 314–326 (1995).
7. McCarthy, G. & Donchin, E. A metric for thought: a comparison of P300 latency and reaction time. *Science* **211**, 77–80 (1981).
8. Coles, M. G. H., Gratton, G. & Donchin, E. Detecting early communication: using measures of movement-related potentials to illuminate human processing. *Biol. Psychol.* **26**, 69–89 (1988).
9. Miller, J. O. & Hackley, S. A. Electrophysiological evidence for temporal overlap among contingent mental processes. *J. Exp. Psychol. Gen.* **121**, 195–209 (1992).
10. van Turennout, M., Hagoort, P. & Brown, C. M. Brain activity during speaking: from syntax to phonology in 40 milliseconds. *Science* **280**, 572–574 (1998).
11. Whalen, P. J. et al. Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *J. Neurosci.* **18**, 411–418 (1998).
12. Morris, J. S., Öhman, A. & Dolan, R. J. Conscious and unconscious emotional learning in the human amygdala. *Nature* **393**, 467–470 (1998).
13. Berns, G. S., Cohen, J. D. & Mintun, M. A. Brain regions responsive to novelty in the absence of awareness. *Science* **276**, 1272–1275 (1997).
14. Sahraie, A. et al. Pattern of neuronal activity associated with conscious and unconscious processing of visual signals. *Proc. Natl Acad. Sci. USA* **94**, 9406–9411 (1997).
15. Tucker, D. Spatial sampling of head electrical fields: the geodesic electrode net. *Electroencephalogr. Clin. Neurophysiol.* **87**, 154–163 (1993).
16. Perrin, F., Pernier, J., Bertrand, D. & Echallier, J. F. Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* **72**, 184–187 (1989).
17. Buckner, R. L. et al. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc. Natl Acad. Sci. USA* **93**, 14878–14883 (1996).

Correspondence and requests for materials should be addressed to S.D. (e-mail: dehaene@shfj.cea.fr).

## Local control of information flow in segmental and ascending collaterals of single afferents

J. Lomelí, J. Quevedo, P. Linares & P. Rudomin

Department of Physiology, Biophysics and Neurosciences, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Apartado Postal 14-740, México D.F. 07000, México

In the vertebrate spinal cord, the activation of GABA( $\gamma$ -aminobutyric acid)-releasing interneurons that synapse with intraspinal terminals of sensory fibres leading into the central nervous system (afferent fibres) produces primary afferent depolarization and presynaptic inhibition<sup>1–3</sup>. It is not known to what extent these



presynaptic mechanisms allow a selective control of information transmitted through specific sets of intraspinal branches of individual afferents<sup>4–7</sup>. Here we study the local nature of the presynaptic control by measuring primary afferent depolarization simultaneously in two intraspinal collaterals of the same muscle spindle afferent. One of these collaterals ends at the L6–L7 segmental level in the intermediate nucleus, and the other ascends to segment L3 within Clarke's column, the site of origin of spinocerebellar neurons<sup>8</sup>. Our results indicate that there are central mechanisms that are able to affect independently the synaptic effectiveness of segmental and ascending collaterals of individual muscle spindle afferents. Focal control of presynaptic inhibition thus allows the intraspinal branches of afferent fibres to function as a dynamic assembly that can be fractionated to convey information to selected neuronal targets. This may be a mechanism by which different spinal postsynaptic targets that are coupled by sensory input from a common source could be uncoupled.

The GABAergic primary afferent depolarization (PAD) reduces the afferent spike amplitude as a result of increased chloride conductance, which leads to reduced neurotransmitter release onto the target neurons<sup>1,2</sup>. In addition, during PAD, the intraspinal spike threshold in afferent fibres in response to externally applied currents is reduced because the resulting depolarization brings regions of the axon close to spike threshold<sup>9</sup>. We used a computer-controlled system<sup>10</sup> to estimate continuously the intraspinal

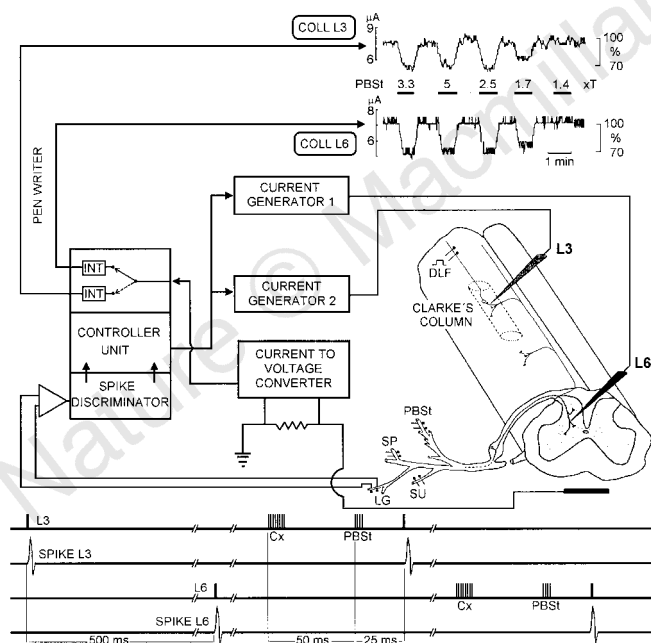
electrical threshold in 20 pairs of L3 and L6 intraspinal collaterals of single afferents from the ankle extensor lateral gastrocnemius nerve (see Fig. 1 and Methods).

We identified afferent fibres from primary muscle spindles by their low peripheral threshold and characteristic PAD pattern, which is different from that usually ascribed to tendon organ<sup>11</sup> and group II (ref. 12) afferents. Stimulation of the knee flexor posterior biceps and semitendinosus (PBSt) nerve reduced the intraspinal threshold in both collaterals of fibres from muscle spindles, because of PAD (Figs 1 and 2c, f). In contrast, for these same fibres, stimulation of cutaneous and articular nerves, or of the motor cortex, bulbar reticular formation and nucleus raphe magnus (NRM), either had no effect on the resting threshold of both collaterals (Fig. 2b;  $n = 15$ ), or increased the spike threshold in at least one of the two collaterals (Fig. 2e,  $n = 5$ ). However, stimulation of any of these latter pathways effectively increased, in one or in both collaterals, the intraspinal threshold already lowered by PBSt stimulation, probably because of inhibition of the evoked PAD (Fig. 2c, f)<sup>13,14</sup>. The increased threshold seen in intraspinal terminals of some muscle spindle afferents in the absence of evoked PAD (Fig. 2e) has been interpreted as due to inhibition of a tonic PAD of GABAergic origin<sup>13</sup> but this does not exclude a possible, albeit unlikely, direct hyperpolarization of the terminals by non-GABAergic mechanisms<sup>15,16</sup>.

For most fibres, the percentage inhibition of the evoked PAD in the two collaterals was of different magnitude, even though in some cases the PBSt-induced PAD in the collaterals was of similar magnitude (Fig. 3a and Table 1). This indicates that the source of variability is the pathways that inhibit the PAD, rather than the pathways that mediate the PAD.

We could modify the relative dominance of inhibition of the evoked PAD in the L3 or L6 collaterals of individual fibres by changing the magnitude of the evoked PAD, as well as by changing the intensity of stimulation of the pathways used to inhibit the PAD, or the location of the cortical stimulus. For 8 out of 20 fibres, we found combinations of stimuli in which conditioning stimuli inhibited almost completely the PAD produced in one collateral, nearly without affecting the PAD produced in the other collateral (Fig. 2f). Differences in the PAD produced in both collaterals, or in the inhibition of the evoked PAD, are expressed by the ratio  $R$ , which is calculated by dividing the percentage threshold changes produced in the L3 collateral by the percentage threshold changes produced in the L6 collateral. The  $R$  values obtained for different fibres during the PBSt-induced PAD showed a relatively low variability (range 0.7–1.1), in contrast with the large variability in the  $R$  values calculated during the inhibition of the evoked PAD (range 0.02–21). We found no significant differences between the mean  $R$  values in the whole set of fibres during inhibition of PAD by a particular input (Table 1). However, in the only case in which we could study three fibres in the same experiment, these afferents had similar  $R$  values in response to the same conditioning stimuli: these  $R$  values were 0.86, 0.90 and 0.85 during PBSt-induced PAD; 2.4, 1.92 and 2.4 during superficial-peroneus-induced inhibition of the PBSt-PAD; 1.36, 1.78 and 1.75 during posterior articular nerve (PAN)-induced inhibition of PAD; and 1.21, 1.99 and 1.99 during cortically induced inhibition of PAD. Therefore, the wide range of differential inhibition of PAD exhibited by the whole set of afferents could be due, at least partly, to differences in the level of tonic descending activity, affecting impulse transmission along the pathways that inhibit PAD in individual experiments<sup>14,17</sup>.

To determine the contribution of descending influences on the differential inhibition of PAD, we used the same set of 20 fibres to study the effects of cold-induced blocking of impulse conduction in the mid-thoracic spinal cord (spinalization). Figure 3 shows the results obtained in one fibre. Before the spinal cold block, conditioning stimulation of the PBSt nerve with trains of pulses 1.3 times threshold reduced the intraspinal threshold of the L3 and L6



**Figure 1** The procedure. Two stimulating micropipettes were inserted into the spinal cord at the L3 and L6 segmental levels (centre right). Current pulses were passed through each micropipette in alternation (from the controller unit to current generator 1 or 2) while intensity was automatically adjusted to produce antidromic responses of a single lateral gastrocnemius (LG) afferent fibre with a constant probability. Current pulses were integrated separately (at INT) and the sampled values maintained until the next test pulse in the series. This allowed a continuous recording of the intraspinal threshold of two collaterals of the same afferent. Top traces show the reduction in threshold due to PAD produced by stimulation of the PBSt nerve with several strengths, expressed as multiples of peripheral threshold (T), as indicated. Calibration bars in this and other figures give the intraspinal threshold in  $\mu$ A and in per cent relative to resting threshold. The same fibre was used here and in Fig. 2d–f. Bottom traces show typical sequences of L3 and L6 threshold testing pulses and of conditioning stimuli. Cx, contralateral motor cortex; SP, superficial peroneus nerve; SU, sural nerve; DLF, dorsolateral funiculus.

**Table 1** Effects of spinalization on strength and inhibition of PAD

|                       |    | Whole sample         | $R < 1$              | $R = 1$             | $R > 1$             |
|-----------------------|----|----------------------|----------------------|---------------------|---------------------|
| PBSt PAD              | R1 | $0.90 \pm 0.11$ (20) | $0.82 \pm 0.09$ (11) | $0.98 \pm 0.02$ (7) | $1.09$ (2)          |
|                       | R2 | $0.94 \pm 0.11$      | $0.88 \pm 0.09$      | $1.01 \pm 0.10$     | 1.01                |
| SU inhibition of PAD  | R1 | $1.9 \pm 2.6$ (18)   | $0.4 \pm 0.3$ (10)   | $1.4$ (1)           | $4.1 \pm 3.0$ (7)   |
|                       | R2 | $2.2 \pm 5.0$        | $3.6 \pm 6.5$        | 0.48                | $0.4 \pm 0.5^{**}$  |
| SP inhibition of PAD  | R1 | $5.5 \pm 12.0$ (18)  | 0.47 (2)             | $1.0 \pm 0.1$ (4)   | $7.9 \pm 14.0$ (12) |
|                       | R2 | $2.2 \pm 4.0$        | 7.38                 | $3.7 \pm 5.7$       | $0.91 \pm 0.8^{**}$ |
| PAN inhibition of PAD | R1 | $1.8 \pm 1.9$ (17)   | $0.5 \pm 0.2$ (6)    | $1.2$ (2)           | $2.9 \pm 2.4$ (9)   |
|                       | R2 | $1.6 \pm 2.0$        | $0.8 \pm 0.5$        | 2.4                 | $1.9 \pm 2.5^{*}$   |
| Cx inhibition of PAD  | R1 | $1.7 \pm 1.6$ (17)   | $0.36 \pm 0.3$ (5)   | $1.2$ (2)           | $2.5 \pm 1.7$ (10)  |

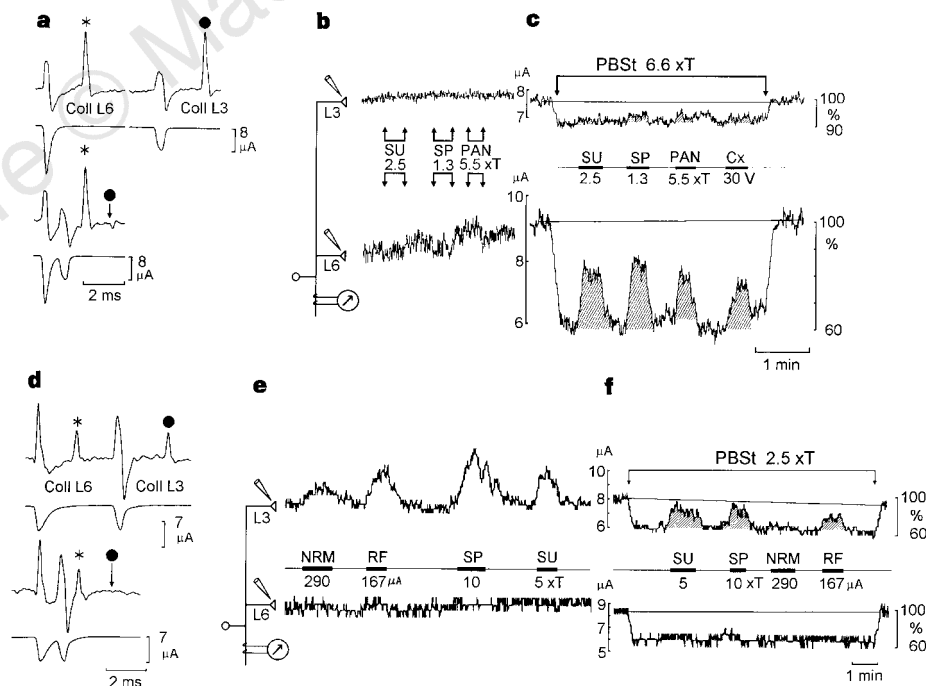
The effects of spinalization on relative strength of PAD and inhibition of PAD in pairs of L3 and L6 collaterals of individual muscle spindle afferents. Columns show means and standard deviations of the  $R$  values, which were obtained by dividing the percentage threshold changes produced by a given procedure in the L3 collateral by the percentage threshold changes produced in the L6 collateral of individual fibres. The last three columns show  $R$  values derived from subsets grouped according to the  $R$  values obtained during pathway stimulation before the spinal block (groups,  $R < 1$ ,  $R = 1$  and  $R > 1$ ). Means and standard deviations were calculated for each category before spinalization (R1) and for the same set of fibres during spinal block (R2). The significance of differences between mean R1 values and R2 values was calculated using the Mann-Whitney test. Double asterisks,  $P < 0.0003$ ; single asterisk,  $0.02 < P < 0.04$ . The numbers of fibres are indicated in parenthesis.

collaterals to 88.4 and 79.5% of control levels, respectively. Conditioning stimulation of the PAN inhibited the PBSt-induced PAD to about the same extent in both collaterals, whereas stimulation of sural and superficial peroneus nerves suppressed almost completely the PAD in the L6 collateral and had only a minor effect on the PAD in the L3 collateral (Fig. 3a). During the cold block, the magnitude of the PBSt-induced PAD in the L3 collateral was about the same as before the block and was slightly reduced in the L6 collateral (thresholds were 86.5 and 88.5% of control, respectively). Nevertheless, the inhibitory effects of sural and superficial peroneus nerves were largely reversed. They now produced a strong inhibition of the PBSt-induced PAD in the L3 collateral, and had a small effect on the PAD in the L6 collateral (Fig. 3b).

Afferent fibres were classified into three categories according to their  $R$  values obtained during pathway stimulation before the spinal block (the categories were  $R < 1$ ,  $R = 1$  and  $R > 1$ ). Means

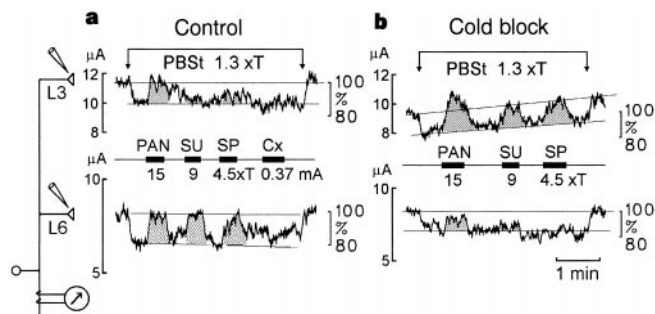
and standard deviations were calculated for each category before (R1) and during (R2) spinalization. The mean  $R$  values for PAD produced by PBSt stimulation were not significantly changed during spinalization (Table 1). However, in those cases in which the inhibition of PAD was larger in the L3 than in the L6 collateral ( $R > 1$ ) before the spinal block, there was a clear and statistically significant reversal in the mean  $R$  values (to  $R < 1$ ) during spinal block. This reversal resulted from a mixture of decreased inhibition in L3 collaterals of some fibres and increased inhibition in L6 collaterals of other fibres.

Our results are evidence for the existence of central mechanisms that allow selection of information flow through segmental and ascending intraspinal collaterals of individual muscle spindle afferents. The segmental information relates to segmental interneurons that mediate a variety of reflex actions (including non-reciprocal postsynaptic inhibition<sup>18</sup> and presynaptic inhibition of muscle

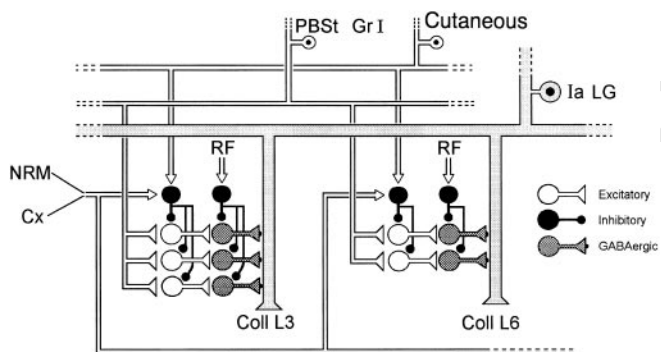


**Figure 2** Selective inhibition of PAD in pairs of collaterals of single afferents. **a**, Upper traces, antidromic responses recorded in an LG fibre evoked by stimulation at L6 (asterisk) and L3 (filled circle) segmental levels; lower traces, stimulating current pulses. At short time intervals between conditioning and testing stimuli, no action potential was recorded in the LG fibre after stimulation of L3, indicating that two collaterals of the same fibre were activated. **b**, Intraspinal threshold changes produced in both collaterals by stimulation of sural (SU), superficial peroneus (SP) and posterior articular (PAN) nerves. **c**, Inhibition of

PBSt-induced PAD by several conditioning stimuli, as indicated. Inhibition of PAD was stronger in the L6 than in the L3 collateral. **d-f**, As for **a-c** but using a different fibre, in which inhibition of evoked PAD was stronger in the L3 than in the L6 collateral. The peripheral thresholds of the two fibres were 1.27 and 1.38 times threshold (T) and the mean conduction velocities were 75.7 and 86.3 m s<sup>-1</sup>. The separation between micropipette tips was 3.3 and 2.95 cm, respectively. RF, bulbar reticular formation.



**Figure 3** Spinalization reverses inhibition of PAD by cutaneous afferents in the L3 and L6 collaterals of same muscle spindle. **a, b**, Inhibition of the PBSt-induced PAD by conditioning stimulation of sural (SU), superficial peroneus (SP) and posterior articular (PAN) nerves, before and during spinal cold block, as indicated. Before the spinal block, conditioning stimulation of the SU and SP nerves produced a stronger inhibition of PAD in the L6 collateral than in the L3 collateral. During spinal block, the inhibition of PAD produced by these same stimuli was stronger in the L3 collateral than in the L6 collateral. Peripheral threshold 1.14 times threshold (T); mean conduction velocity  $84.3 \text{ m s}^{-1}$ . Separation between micropipette tips, 2.5 cm.



**Figure 4** Neuronal connections explaining the local character of inhibition of PAD in the L3 and L6 collaterals of individual muscle spindle afferents. PAD produced by stimulation of group I PBSt afferents is mediated by at least two interposed interneurons<sup>29</sup>. Separate groups of GABAergic interneurons produce PAD of L3 and L6 collaterals. Stimulation of cutaneous afferents and of raphe-spinal (NRM) and corticospinal (Cx) fibres inhibits PAD by acting on the first-order interneurons mediating PAD of muscle spindles, through separate sets of inhibitory interneurons. Reticulospinal (RF) fibres reduce PAD by inhibiting the last-order GABAergic interneurons<sup>30</sup>.

spindles and tendon organs<sup>19</sup>), whereas the ascending information relates to spinocerebellar neurons that relay information from muscle afferents to the cerebellum<sup>8</sup>, and is of relevance for sensory acquisition and discrimination<sup>20</sup> and for motor learning<sup>21</sup>. This local control of information transmission seems to be possible because separate sets of interneurons mediate PAD of segmental and ascending branches of individual muscle spindle afferents<sup>22,23</sup> and, as shown here, because descending activity affects specific sets of PAD-inhibiting interneurons (Fig. 4). The extent to which the differential presynaptic control in single fibres is reflected at the population level will depend on the degree of synchronization of the PAD-mediating interneurons, a feature that is also centrally controlled<sup>19,24</sup>.

Mechanisms similar to those described here could be involved in the changes in the presynaptic inhibition of muscle spindle afferents that occur in humans at the onset of a voluntary contraction<sup>25</sup>, or in the reduction of presynaptic inhibition that follows mechanical activation of cutaneous afferents<sup>26</sup>. Our results support the conclusion that descending fibre systems contribute to the 'selection' of the

kind of afferent information that is needed for the execution of specific motor tasks<sup>27</sup>. Descending fibres in higher vertebrates seem to be particularly suited for this selection process, because their own synaptic effectiveness is not subject to the GABAergic presynaptic control mechanisms present in sensory fibres<sup>9,28</sup>. Focal control of presynaptic inhibition in axonal terminals has also been observed in reticulospinal neurons in the lamprey and in the nervous system of invertebrates<sup>2</sup>. Thus, functional compartmentalization of neuronal output can be a general mechanism that is used to convey information to selected neuronal targets. □

## Methods

**Preparation.** Guidelines contained in NIH publication 85-13, revised in 1985, on the principles of laboratory animal care were followed throughout. Briefly, we studied, in 11 experiments, adult cats initially anaesthetized with pentobarbitone (40 mg per kg weight, intraperitoneally), supplemented during the dissection with additional doses of  $10 \text{ mg kg}^{-1}$  intravenously, as necessary to maintain deep anaesthesia. After the surgical procedures the animals were paralysed and artificially respired. Deep anaesthesia was maintained during the experiment by intravenous injections of pentobarbitone ( $3 \text{ mg h}^{-1}$ ). Adequacy of anaesthesia was assessed by verifying that the pupils were constricted and that blood pressure was stable (100–120 mm Hg) and not affected by noxious stimulation to the skin. The left sural, superficial peroneus, posterior articular, and posterior biceps and semitendinosus nerves were sectioned and prepared for stimulation. The medial and lateral gastrocnemius nerves were also sectioned, and several (10–15) fine filaments of the central end of the lateral gastrocnemius nerve were dissected and mounted for recording (Fig. 1). The lumbosacral and low thoracic spinal segments were exposed and the left L5 to S1 ventral roots were sectioned. A frontal craniotomy was performed to expose the contralateral motor cortex. In some experiments the floor of the fourth ventricle was exposed to stimulate the ipsilateral nucleus raphe magnus and the bulbar (magnocellular) reticular formation. Exposed tissues were covered with mineral oil to prevent desiccation and kept at constant temperature ( $37^\circ\text{C}$ ).

**Electrode placement.** One glass micropipette (filled with 2 M NaCl, 1.2–1.8 MΩ) was inserted within the intermediate nucleus at segment L6–L7 using as a guide the extracellular field potentials produced by stimulation of the lateral gastrocnemius and the PBSt nerves<sup>4,5</sup>. A second micropipette was inserted within Clarke's column at the L3 segment, aided by recording antidromic field potentials produced by stimulation of the dorsal spinocervellar tract in the dorsolateral funiculus<sup>22,23</sup>. Once in place, stimulating pulses ( $400\text{--}\mu\text{s}$  duration,  $2\text{--}12 \mu\text{A}$ ) were passed through each micropipette, through independent current pulse generators (Fig. 1) and stimulating micropipettes were displaced until both of them produced antidromic responses of a single fibre in the same lateral gastrocnemius nerve filament. Interaction by refractoriness between the antidromic responses produced by both electrodes was taken as evidence that these responses resulted from activation of two collaterals of the same afferent fibre<sup>4,5</sup> (Fig. 2a,d). The selected afferents had conduction velocities between 62 and  $97 \text{ m s}^{-1}$  (mean  $75.0 \pm 11.8 \text{ m s}^{-1}$ ) and peripheral thresholds of 1.01–1.52 times the threshold of the most excitable fibres (mean  $1.27 \pm 1.2$  times threshold).

**Intraspinal threshold measurements.** Separate stimulating current pulses were applied in alternation (at 2 Hz) through each micropipette. Afferent spikes were passed through a window discriminator. Stimulus intensity was independently adjusted, by means of a digital controller unit, to produce, in each case, antidromic firing with a constant probability (set to 0.5) that was continuously calculated using an exponentially weighted estimate (in the past) with a time constant of about ten successive events<sup>10</sup>. The current pulses were measured in the return path to ground, and were separately integrated and recorded in different channels of a penwriter<sup>10</sup> (Fig. 1). Changes in PAD were estimated from the intraspinal threshold changes, measured as a percentage change relative to the resting threshold, once a steady value was attained. These changes do not seem to depend on the position of the micropipette relative to the tested afferent fibre (Fig. 1 in ref. 5) and are taken to represent events occurring in the terminals themselves. Although shifted by one cycle, threshold changes in the L3 and L6 collaterals can be considered as simultaneous, because they are measured at steady state. The PBSt nerve was stimulated to produce



PAD with a train of four pulses at 300 Hz (1.25–6.6 times threshold), applied 25 ms before each intraspinal threshold testing pulse. To inhibit the PBSt-induced PAD, the superficial peroneus and the posterior articular nerves were stimulated with single pulses (1.3–15 times threshold), and the motor cortex (200–500  $\mu$ A), nucleus raphe magnus and reticular formation (50–100  $\mu$ A) were stimulated with eight pulses at 700 Hz, all applied 75 ms before the PBSt train (low traces in Fig. 1).

**Spinalization.** Spinalization was produced by blocking impulse conduction by means of a cooled probe placed over the thoracic spinal cord. The degree of spinal block was assessed by recording the descending volleys produced by stimulation of the motor cortex, bulbar reticular formation or nucleus raphe magnus<sup>14</sup>.

Received 9 January; accepted 5 August 1998.

- Rudomin, P. in *The Segmental Motor System* (eds Binder, M.S. & Mendell, L. M.) 349–380 (Oxford Univ. Press, New York, 1990).
- Nusbaum, M. P., El-Manira, A., Gossard, J. P. & Rossignol, S. in *Neurons, Networks and Motor Behaviour* (eds Stein, P. S. G., Grillner, S., Selverston, A. I. & Stuart, D. G.) 237–253 (MIT Press, Cambridge, Massachusetts, 1997).
- Gossard, J. P. Control of transmission in muscle Ia afferents during fictive locomotion in cat. *J. Neurophysiol.* **76**, 4104–4112 (1996).
- Quevedo, J., Eguibar, J. R., Lomeli, J. & Rudomin, P. Patterns of connectivity of spinal interneurons with single muscle afferents. *Exp. Brain Res.* **115**, 387–402 (1997).
- Eguibar, J. R., Quevedo, J. & Rudomin, P. Selective cortical and segmental control of primary afferent depolarization of single muscle afferents in the cat spinal cord. *Exp. Brain Res.* **113**, 411–430 (1997).
- Luscher, H. R. in *Presynaptic Inhibition and Neural Control* (eds Rudomin, P., Romo, R. & Mendell, L. M.) 126–137 (Oxford Univ. Press, Oxford, 1998).
- Wall, P. D. Do nerve impulses penetrate terminal arborizations? *Trends Neurosci.* **18**, 99–103 (1995).
- Mann, M. D. Clarke's column and the dorsal spinocerebellar tract. *Brain Behav. Evol.* **7**, 34–83 (1973).
- Rudomin, P., Engberg, I. & Jiménez, I. Mechanisms involved in presynaptic depolarization of group I and rubrospinal fibers in cat spinal cord. *J. Neurophysiol.* **72**, 532–548 (1981).
- Madrid, J., Alvarado, J., Dutton, H. & Rudomin, P. A method for the dynamic continuous estimation of excitability changes of single fiber terminals in the central nervous system. *Neurosci. Lett.* **11**, 253–258 (1979).
- Enriquez, M., Jiménez, I. & Rudomin, P. Segmental and supraspinal control of synaptic effectiveness of functionally identified muscle afferents in the cat. *Exp. Brain Res.* **107**, 391–404 (1996).
- Riddell, J. S., Jankowska, A. & Eide, E. Depolarization of group II muscle afferents by stimuli applied in the locus coeruleus and raphe nuclei of the cat. *J. Physiol. (Lond)* **461**, 723–741 (1993).
- Rudomin, P., Núñez, R., Madrid, J. & Burke, R. E. Primary afferent hyperpolarization and presynaptic facilitation of Ia afferent terminals induced by large cutaneous fibres. *J. Neurophysiol.* **37**, 413–429 (1979).
- Quevedo, J., Eguibar, J. R., Jiménez, I., Schmidt, R. F. & Rudomin, P. Primary afferent depolarization of muscle afferents elicited by stimulation of joint afferents in cats with intact neuaxis and during reversible spinalization. *J. Neurophysiol.* **70**, 1–12 (1993).
- Chan, S. H. H. & Barnes, C. D. A presynaptic mechanism evoked from brainstem reticular formation in the lumbar cord and its temporal significance. *Brain Res.* **45**, 101–114 (1972).
- Curtis, D. R., Leah, J. D. & Peet, M. J. Effects of noradrenaline and 5-hydroxytryptamine on spinal Ia afferent terminations. *Brain Res.* **258**, 328–332 (1983).
- Lundberg, A. in *Physiology of Spinal Nerve* (eds Eccles, J. C. & Schädé, J. P.) 197–221 (Elsevier, Amsterdam, 1964).
- Jankowska, E. Interneuronal relay in pathways from proprioceptors. *Prog. Neurobiol.* **38**, 335–378 (1992).
- Rudomin, P., Solodkin, M. & Jiménez, I. Synaptic potentials of primary afferent fibers and motoneurons evoked by single intermediate nucleus interneurons in the cat spinal cord. *J. Neurophysiol.* **57**, 1288–1313 (1987).
- Geo, J. H. et al. Cerebellum implicated in sensory acquisition and discrimination rather than motor control. *Science* **272**, 545–547 (1996).
- Raymond, J. L., Lisberger, S. G. & Mauk, M. D. The cerebellum: a neuronal learning machine? *Science* **272**, 1126–1131 (1996).
- Jankowska, E. & Padel, Y. On the origin of presynaptic depolarization of group I muscle afferents in Clarke's column in the cat. *Brain Res.* **295**, 195–201 (1984).
- Harrison, P. J. & Jankowska, E. Do interneurons in the lower lumbar segments contribute to presynaptic depolarization of group I muscle afferents in Clarke's column? *Brain Res.* **295**, 203–210 (1984).
- Rudomin, P., Burke, R. E., Núñez, R., Madrid, J. & Dutton, H. Control by presynaptic correlation: a mechanism affecting information transmission from Ia fibers to motoneurons. *J. Neurophysiol.* **38**, 267–284 (1975).
- Hultborn, H., Meunier, S., Morin, C., Pierrot-Deseilligny, E. & Shindo, E. Changes in presynaptic inhibition of Ia fibres at the onset of voluntary contraction of homonymous and synergistic muscles in man. *J. Physiol. (Lond.)* **389**, 757–772 (1987).
- Iles, J. F. Evidence for cutaneous and corticospinal modulation of presynaptic inhibition of Ia afferents from the human lower limb. *J. Physiol. (Lond.)* **491**, 197–207 (1996).
- Nelson, R. J. Interactions between motor commands and somatic perception in sensorimotor cortex. *Curr. Opin. Neurobiol.* **6**, 801–810 (1996).
- Curtis, D. R., Wilson, W. J. & Malik, R. The effect of GABA on the terminations of vestibulospinal neurons in the cat spinal cord. *Brain Res.* **291**, 372–375 (1984).
- Jankowska, E., McCrea, D., Rudomin, P. & Sykova, E. Observations on neuronal pathways subserving primary afferent depolarization. *J. Neurophysiol.* **46**, 506–516 (1981).
- Rudomin, P., Jiménez, I., Solodkin, M. & Duénas, S. Sites of action of segmental and descending control of transmission on pathways mediating PAD of Ia- and Ib-afferent fibers in cat spinal cord. *J. Neurophysiol.* **50**, 743–769 (1983).

**Acknowledgements.** We thank C. León for help with the experiments and J. E. Velazquez, P. Reyes and L. Jiménez for computational support. This study was supported partly by grants from the NIH and CONACyT.

Correspondence and requests for materials should be addressed to R.P. (e-mail: rudomin@fisio.cinvestav.mx).

## Drosophila Tcf and Groucho interact to repress Wingless signalling activity

Robert A. Cavallo<sup>\*†</sup>, Rachel T. Cox<sup>\*†</sup>, Melissa M. Molíne<sup>‡</sup>, Jeroen Roose<sup>§</sup>, Gordon A. Polevoy<sup>\*</sup>, Hans Clevers<sup>§</sup>, Mark Peifer<sup>\*</sup> & Amy Bejsovec<sup>‡</sup>

<sup>\*</sup> Department of Biology and Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA

<sup>‡</sup> Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208, USA

<sup>§</sup> Department of Immunology, University Hospital, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

<sup>†</sup> These authors contributed equally to this work

Wingless/Wnt signalling directs cell-fate choices during embryonic development<sup>1,2</sup>. Inappropriate reactivation of the pathway causes cancer<sup>3–5</sup>. In *Drosophila*, signal transduction from Wingless stabilizes cytosolic Armadillo<sup>1</sup>, which then forms a bipartite transcription factor with the HMG-box protein *Drosophila* Tcf (dTcf) and activates expression of Wingless-responsive genes<sup>6–8</sup>. Here we report that in the absence of Armadillo, dTcf acts as a transcriptional repressor of Wingless-responsive genes, and we show that Groucho acts as a corepressor in this process. Reduction of dTcf activity partially suppresses *wingless* and *armadillo* mutant phenotypes, leading to derepression of Wingless-responsive genes. Furthermore, overexpression of wild-type dTcf enhances the phenotype of a weak *wingless* allele. Finally, mutations in the *Drosophila* *groucho* gene also suppress *wingless* and *armadillo* mutant phenotypes as Groucho physically interacts with dTcf and is required for its full repressor activity.

Mutations of Tcf-binding sites in the promoters of *Drosophila*

**Table 1 Genetic interactions between *arm*, *dTcf* and *gro***

| Genetic cross                                                                                                | Per cent in weakest classes |
|--------------------------------------------------------------------------------------------------------------|-----------------------------|
| <i>arm</i> <sup>XP33</sup> /FM7 × +/+                                                                        | 1 (n = 402)                 |
| <i>arm</i> <sup>XP33</sup> /FM7 × Df(4)M62f/+                                                                | 21 (n = 559)                |
| <i>arm</i> <sup>XP33</sup> /FM7 × Df(4)M63a/+                                                                | 0 (n = 691)                 |
| <i>arm</i> <sup>XP33</sup> /FM7 × <i>ci</i> <sup>9</sup> /+                                                  | 13 (n = 318)                |
| <i>arm</i> <sup>XP33</sup> /FM7 × <i>ci</i> <sup>9ell</sup> /+                                               | 2 (n = 353)                 |
| <i>arm</i> <sup>XP33</sup> /FM7 × dTCF <sup>1</sup> /+                                                       | 15 (n = 350)                |
| <i>arm</i> <sup>XP33</sup> /FM7 × dTCF <sup>2</sup> /+                                                       | 27 (n = 299)                |
| <i>arm</i> <sup>XP33</sup> /FM7 × dTCF <sup>3</sup> /+                                                       | 23 (n = 595)                |
| <i>arm</i> <sup>XP33</sup> /FM7 × <i>gro</i> <sup>BX22</sup> /+                                              | 0 (n = 159)                 |
| <i>arm</i> <sup>XP33</sup> /FM7 × <i>gro</i> <sup>E48</sup> /+                                               | 1 (n = 151)                 |
| <i>arm</i> <sup>XP33</sup> /+; <i>gro</i> <sup>BX22</sup> /+ × <i>gro</i> <sup>BX22</sup> /+                 | 8 (n = 121)                 |
| <i>arm</i> <sup>XP33</sup> /+; <i>gro</i> <sup>E48</sup> /+ × <i>gro</i> <sup>E48</sup> /+                   | 12 (n = 378)                |
| <i>arm</i> <sup>XP33</sup> /+; <i>gro</i> <sup>BX22</sup> /+ × <i>arm</i> <sup>XP33</sup> /y <sup>arm+</sup> | 10 (n = 132)                |
| <i>arm</i> <sup>YD35</sup> /FM7 × +/+                                                                        | 14 (n = 171)                |
| <i>arm</i> <sup>YD35</sup> /FM7 × dTCF <sup>2</sup> /+                                                       | 46 (n = 321)                |
| <i>arm</i> <sup>YD35</sup> /FM7 × dTCF <sup>3</sup> /+                                                       | 42 (n = 353)                |
| <i>arm</i> <sup>YD35</sup> /FM7 × <i>gro</i> <sup>BX22</sup> /+                                              | 10 (n = 169)                |
| <i>arm</i> <sup>YD35</sup> /FM7 × <i>gro</i> <sup>E48</sup> /+                                               | 17 (n = 196)                |
| <i>arm</i> <sup>YD35</sup> /+; <i>gro</i> <sup>BX22</sup> /+ × <i>gro</i> <sup>BX22</sup> /+                 | 46 (n = 183)                |
| <i>arm</i> <sup>YD35</sup> /+; <i>gro</i> <sup>E48</sup> /+ × <i>gro</i> <sup>E48</sup> /+                   | 35 (n = 169)                |
| <i>arm</i> <sup>XM19</sup> /FM7 × +/+                                                                        | 3 (n = 214)                 |
| <i>arm</i> <sup>XM19</sup> /FM7 × dTCF <sup>2</sup> /+                                                       | 46 (n = 291)                |
| <i>arm</i> <sup>XM19</sup> /FM7 × dTCF <sup>3</sup> /+                                                       | 43 (n = 305)                |
| <i>wg</i> <sup>CX4</sup> /CyO × UAS-dTCF/+; Df(2)DE/+                                                        | 99 (n = 216)                |
| <i>wg</i> <sup>CX4</sup> E22C/+ × UAS-dTCF/+; Df(2)DE/+                                                      | 47 (n = 308)                |
| <i>wg</i> <sup>CX4</sup> E22C/+; <i>gro</i> <sup>BX22</sup> /+ male × UAS-dTCF/+; Df(2)DE/+                  | 49 (n = 415)                |
| <i>wg</i> <sup>CX4</sup> E22C/+; <i>gro</i> <sup>BX22</sup> /+ female × UAS-dTCF/+; Df(2)DE/+                | 94 (n = 223)                |
| E22C-GAL4/+ × UAS-dTCF-ΔN (line 5)                                                                           | 3 (n = 259)                 |
| <i>gro</i> <sup>E48</sup> /+; E22C-GAL4/+ × dTCF-ΔN (line 5)                                                 | 97 (n = 206)                |
| E22C-GAL4/+ × dTCF-ΔN (line 1)                                                                               | 3 (n = 216)                 |
| <i>gro</i> <sup>E48</sup> /+; E22C-GAL4/+ × dTCF-ΔN (line 1)                                                 | 93 (n = 231)                |

Outlines were scored using the criteria in ref. 15. In each case, we calculated the percentage of embryos in the two least severe phenotypic categories.

*Ultrabithorax* (*Ubx*)<sup>7</sup> or *Xenopus siamensis*<sup>9</sup> reduce the level of gene expression in the normal expression domain of the animal, as expected for perturbation of a transcriptional activator. Surprisingly, however, these mutations also result in ectopic gene expression outside the normal domain. This led to the proposal that Tcf proteins may act as repressors<sup>7,9</sup>. We have tested this hypothesis and found that dTcf can function as either an activator or a repressor of Wingless (Wg)-responsive genes depending on the state of the Wg signalling pathway and thus the availability of Armadillo (Arm), dTcf's coactivator.

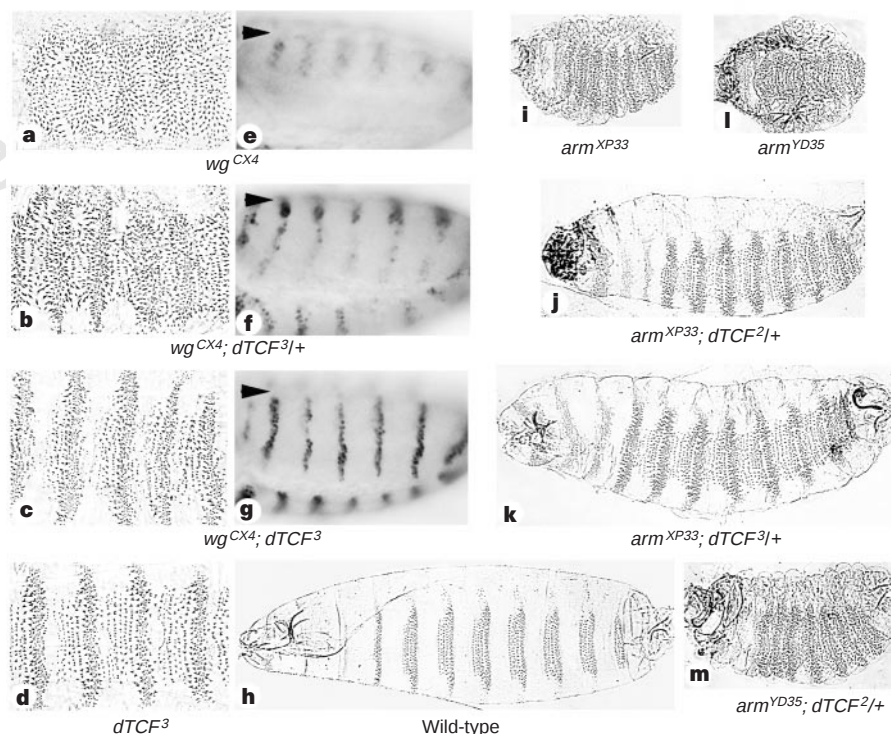
Reducing the level of zygotic dTcf, by making embryos heterozygous for a null *dTcf* mutation<sup>8</sup>, suppresses the segment polarity phenotype that is a consequence of a *wg* null allele (Fig. 1a, b), consistent with a repressive function for dTcf. Indeed, the previously described *dTcf* null mutations<sup>8</sup> were isolated as dominant suppressors of *wg*<sup>11,14</sup> (A.B., unpublished observations). Although the function of maternal dTcf has not been determined, we saw no difference in suppression of *wg* mutations when the mutant *dTcf* allele was derived from the mother or the father. This indicates that little, if any, maternal dTcf participates in the repressive effect. Complete loss of zygotic dTcf suppresses the *wg* cuticle pattern defect even more substantially (Fig. 1c), *wg*; *dTcf* double mutant embryos are larger and exhibit a greater variety of cell types. The double mutant phenotype closely resembles the *dTcf* zygotic null phenotype (Fig. 1d), supporting the idea that the *dTcf* single mutant lacks both repressor and activator function and therefore should be insensitive to the removal of Wg, the upstream activator.

To determine whether, in the absence of Wg activity, dTcf normally represses Wg target genes, we studied the Wg-responsive gene *engrailed* (*en*), which is expressed in epidermal cells just posterior to the *wg*-expressing row<sup>10–12</sup> and is dependent on Wg activity for maintenance of expression<sup>13,14</sup>. *wg* null mutants com-

pletely lose epidermal *en* expression before stage 10 (Fig. 1e), but in homozygous *wg* embryos that are heterozygous for *dTcf*, some cells maintain *en* expression (Fig. 1f). Homozygous *wg*; *dTcf* mutants maintain more *en* expression than heterozygous mutants (Fig. 1g), similar to that observed in *dTcf* zygotic null mutants<sup>8</sup>. This corroborates a repressive role for dTcf in cells in which the Wg signalling pathway is not active. Thus the severe phenotype of a *wg* null mutant reflects the loss of activation by Arm–dTcf while repression by dTcf remains intact. In contrast, both activation and repression are affected in a *dTcf* zygotic null embryo, resulting in less severe disruption of the cuticle pattern and Wg-responsive gene expression.

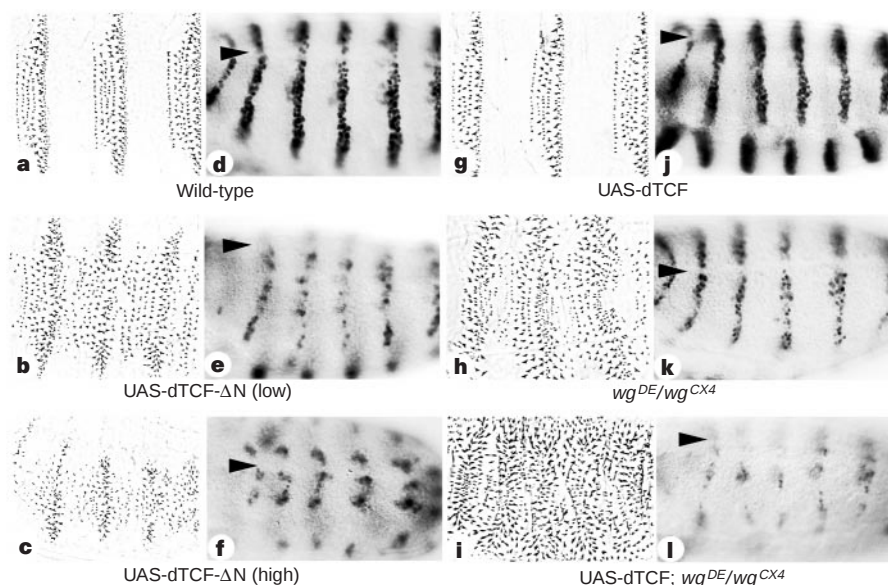
We further established that a reduction in Arm levels causes dTcf to act as a repressor. A dose-sensitive screen for suppressors of *arm*<sup>XP33</sup> (R.T.C. and M.P., unpublished observations) revealed interactions of *arm*<sup>XP33</sup> with *dTcf* that were similar to those between *wg* and *dTcf*. Embryos zygotically mutant for *arm*<sup>XP33</sup> show a strong polarity phenotype<sup>15</sup>, characterized by reduced size, a lawn of denticles and missing head structures (Fig. 1h, i). Heterozygosity for the entire fourth chromosome or for the chromosomal deletion *Df(4)M62f* suppresses *arm* mutations whereas heterozygosity for *Df(4)M63a* does not. Two genes in the suppressing region, *ci* and *dTcf*, are required for normal segment polarity; *ci* acts in Hedgehog signalling<sup>16,17</sup>. Two *dTcf* null alleles strongly suppress the *arm*<sup>XP33</sup> mutation (Fig. 1j, k), as does *ci*<sup>D</sup>, a *ci* *dTcf* double mutant, whereas *ci*<sup>Cell</sup>, a putative *ci* null, does not. Suppression is not allele-specific; *dTcf* suppresses the zygotic null *arm*<sup>YD35</sup> (Fig. 1l, m) and the moderate hypomorph *arm*<sup>XM19</sup> (Table 1).

We previously constructed a dTcf molecule (dTcf-ΔN), lacking the putative Arm-binding domain, that antagonizes Wg signalling when ubiquitously expressed during embryogenesis<sup>8</sup>. The antagonism depends on expression level, ranging from a weak phenotype,



**Figure 1** *dTcf* is a dose-dependent suppressor of *wg* and *arm*. In all photographs, anterior is to the left. **a**, *wg*<sup>CX4</sup> null homozygote. **b**, *wg*<sup>CX4</sup>; *dTcf*<sup>3</sup>/+ embryos show partial suppression of the *wg* phenotype (*n* > 100). **c**, *wg*<sup>CX4</sup>; *dTcf*<sup>3</sup> embryos show more extensive suppression (*n* > 300). **d**, *dTcf*<sup>3</sup> homozygotes resemble *wg*<sup>CX4</sup>; *dTcf*<sup>3</sup>. **e**, *wg*<sup>CX4</sup> mutant embryos lose epidermal En antibody staining (En staining in nervous system is below the plane of focus). The arrowhead indicates

the posterior ventral midline. **f**, *wg*<sup>CX4</sup>; *dTcf*<sup>3</sup>/+ embryos retain some En-staining cells (*n* > 100). **g**, En is further derepressed in *wg*<sup>CX4</sup>; *dTcf*<sup>3</sup> double homozygotes (*n* > 200). **h**, Wild-type embryo (**h–m** are at the same magnification, lower than that in **a–g**). **i**, *arm*<sup>XP33</sup>/Y. **j**, *arm*<sup>XP33</sup>/Y; *dTcf*<sup>2</sup>/+, and **k**, *arm*<sup>XP33</sup>/Y; *dTcf*<sup>3</sup>/+ show partial suppression of the *arm*<sup>XP33</sup>/Y phenotype. **l**, *arm*<sup>YD35</sup>/Y. **m**, *arm*<sup>YD35</sup>/Y; *dTcf*<sup>2</sup>/+ shows substantial suppression of the *arm*<sup>YD35</sup>/Y phenotype.



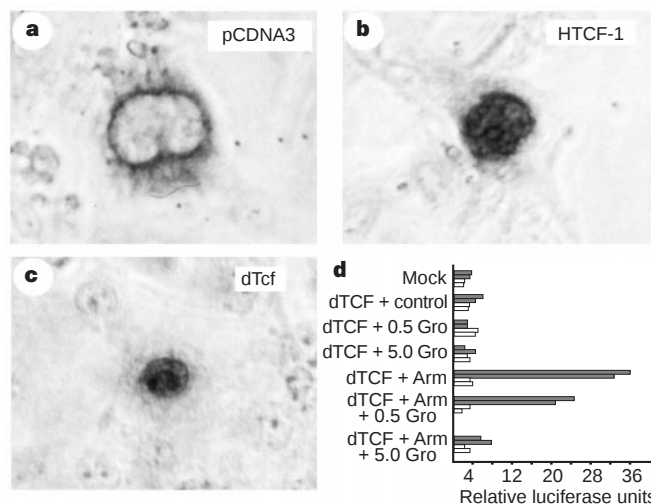
**Figure 2** dTcf represses Wg-responsive genes. **a–f**, dTcf-ΔN is a constitutive repressor. **a**, A UAS-dTcf-ΔN construct without a Gal4 driver produces a wild-type phenotype. UAS-dTcf-ΔN expressed ubiquitously with E22C-GAL4 produces **b**, moderate to **c**, severe segment polarity phenotypes. **d**, Normal En antibody staining in a stage 10 embryo containing UAS-dTcf-ΔN without a Gal4 driver. The arrowhead indicates the posterior ventral midline. UAS-dTcf-ΔN expressed ubiquitously reduces **(e)** or eliminates **(f)** epidermal *en* expression. **g–l**, Wild-

type dTcf acts as a repressor when Wg signalling is limited. **g**, Ubiquitous UAS-dTcf changes the wild-type cuticle pattern only subtly. **h**, *Df(2)DE/wg<sup>CX4</sup>* embryos show a weak *wg* phenotype. **i**, Ubiquitous UAS-dTcf expression in *Df(2)DE/wg<sup>CX4</sup>* embryos results in a severe *wg* phenotype. **j**, Ubiquitous UAS-dTcf does not disrupt wild-type *en* expression. **k**, *Df(2)DE/wg<sup>CX4</sup>* embryos show slightly reduced En staining. **l**, En staining in *Df(2)DE/wg<sup>CX4</sup>* embryos is repressed by ubiquitous UAS-dTcf.

similar to that produced by the *dTcf* zygotic null allele (Fig. 2a, b), to a strong phenotype resembling that resulting from a *wg* null allele (Fig. 2c). Ubiquitous dTcf-ΔN expression also reduces *en* expression; the normal epidermal En stripe (Fig. 2d) is reduced to scattered *en*-expressing cells (Fig. 2e) or is completely repressed (Fig. 2f), mimicking *wg* loss of function. Full-length dTcf does not have this effect (Fig. 2g). Thus, dTcf-ΔN, lacking the Arm-binding region, acts as a constitutive repressor. This reconciles our results with those of ref. 6, in which *dTcf* mutations were isolated as suppressors of *wg* hyperactivity. These *dTcf* alleles contained amino-terminal missense mutations which reduced Arm-binding<sup>6</sup>, and, thus, should selectively disrupt dTcf's activation function but leave intact its repressive function.

Thus the difference between dTcf in its role as activator versus repressor seems to reflect a balance between dTcf with and without Arm. Overexpression of full-length dTcf in a normal embryo does not antagonize Wg signalling (Fig. 2g, j). However, when levels of Arm are reduced by limiting Wg activity, similar overexpression of full-length dTcf represses Wg target genes. We lowered Wg activity by using *Df(2)DE*<sup>18</sup>, which removes part of the *wg* regulatory region (A.B., data not shown); Wg signalling is reduced but still specifies many wild-type pattern elements (Fig. 2h) and stabilizes some epidermal *en* expression (Fig. 2k). Overexpression of full-length dTcf in these embryos both eliminates wild-type pattern elements (Fig. 2j) and represses *en* expression (Fig. 2l). As overexpression of dTcf has no effect on *en* expression in wild-type embryos (Fig. 2j), we conclude that whether dTcf acts as an activator or a repressor depends on the level of Wg signalling and probably on the amount of available Arm.

In cultured mammalian cells, dTcf alone does not repress transcription of reporter genes<sup>8</sup>. Thus, it seemed likely that dTcf requires a corepressor, as it requires Arm as a coactivator. We have found that Tcf-1 binds to vertebrate Grg family members, homologues of *Drosophila* Groucho (Gro), a known corepressor<sup>19</sup>. We therefore tested whether *Drosophila* Gro binds dTcf (Fig. 3). When we expressed the N-terminal region of *Drosophila* Gro (amino acids 1–181) in COS cells, it localized to the cytoplasm (Fig. 3a).



**Figure 3** dTcf interacts with *Drosophila* Gro. **a**, A Myc-tagged N-terminal fragment of Gro protein (amino acids 1–181) localizes to the cytoplasm of COS cells. Co-transfection with **b**, human Tcf-1 or **c**, *Drosophila* Tcf results in nuclear localization of Gro(1–181). **d**, Arm-dTcf-mediated transactivation of a dTcf reporter gene is repressed by Gro. IIA1.6 B cells were transfected with the indicated expression plasmids. Shaded bars indicate luciferase activity of pTKTOP, a reporter plasmid containing three optimal dTcf-binding sites upstream of the minimal HSV-TK promoter; unshaded bars indicate activity of pTKFOP, a similar construct with mutated dTcf-binding sites.

Coexpression of either human Tcf-1 (Fig. 3b) or dTcf (Fig. 3c) results in the localization of Gro(1–181) to the nucleus, consistent with a physical association between the proteins. Full-length Gro is constitutively nuclear (data not shown), and, as such, is not informative in this assay. The nuclear recruitment of Gro(1–181) by dTcf is very similar to the recruitment of  $\beta$ -catenin<sup>20</sup>, a known Tcf-binding partner. We then determined whether the dTcf–Gro



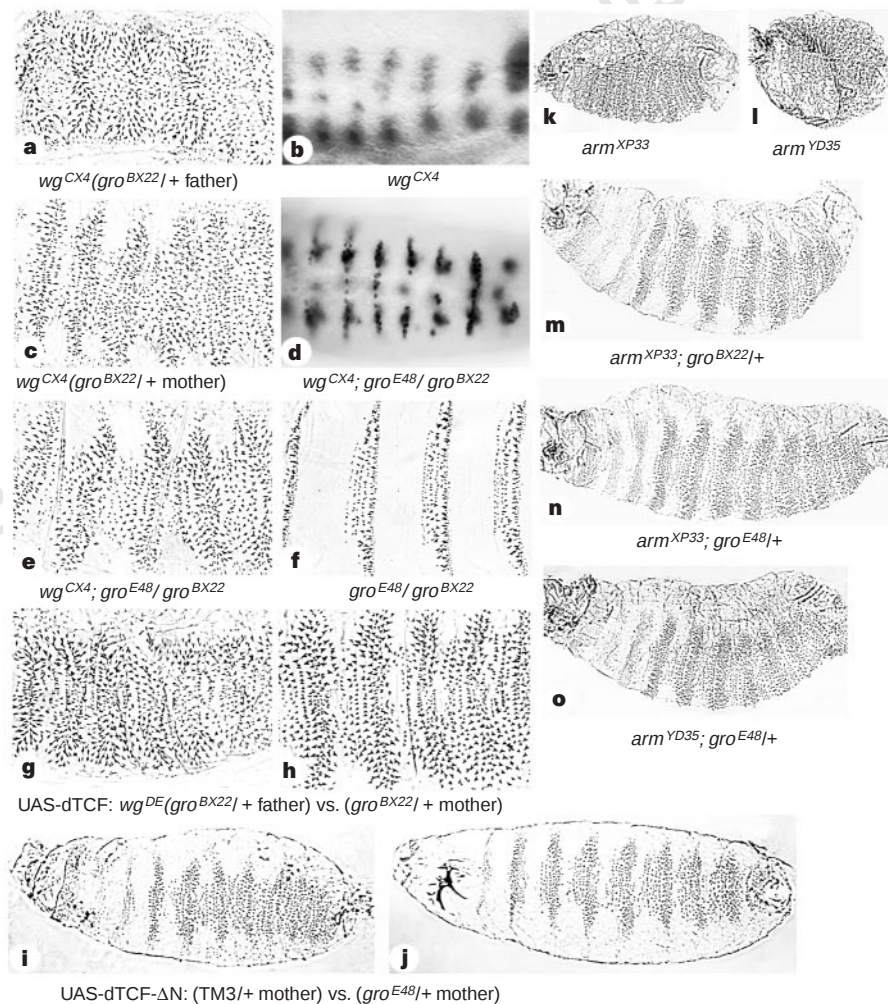
association has functional consequences (Fig. 3d). We transfected IIA1.6 B cells with reporter genes containing dTcf-binding sites upstream of the thymidine kinase (TK) promoter. Expression of dTcf alone leads to very little reporter activation, whereas coexpression of dTcf and Arm leads to robust activation<sup>8</sup>. Expressing *Drosophila* Gro results in a dose-sensitive reversal of coactivation (Fig. 3d), again consistent with a physical interaction.

To determine if Gro acts as a corepressor *in vivo*, we tested whether *gro* mutants interact genetically with the Wg signalling pathway. *Drosophila* Gro acts as a corepressor for several transcription factors and is essential for dosage compensation, early segmentation and neurogenesis<sup>21</sup>. These pleiotropic effects would obscure later effects on epidermal patterning by the Wg pathway. However, as with *dTcf*, we find that *gro* mutations show dose-sensitive interactions with both *wg* and *arm*. Reducing the dose of maternal Gro suppresses the *wg* null phenotype, whereas reduction of paternal Gro has no effect (Fig. 4a, c). Zygotic reduction does not substantially increase rescue of the *wg* null phenotype (Fig. 4e). Suppression of this phenotype correlates with partial derepression of epidermal *en* expression (Fig. 4b, d). Reduction of both maternal

and zygotic Gro also strongly suppresses *arm* phenotypes (Fig. 4k–o), whereas reduction of zygotic Gro alone does not. Two other aspects of the *gro* phenotype support a function for Gro in Wg signalling. First, some zygotically mutant *gro* embryos retain sufficient ventral epidermis to secrete a cuticle pattern exhibiting subtle patterning defects, consistent with mild Wg hyperactivation (Fig. 4f). Second, *gro* zygotic mutants show expanded *en* expression<sup>22</sup>, as seen when Wg is hyperactivated<sup>23</sup>.

If dTcf and Gro act together to mediate repression, dTcf repression should require Gro. We thus tested whether a reduction in Gro dosage diminishes dTcf's effectiveness as a repressor. Excess dTcf represses Wg target genes in weak *wg* mutant embryos (Fig. 2i, k, l). However, when maternal Gro levels are reduced, dTcf repression is decreased significantly (Fig. 4g, h). Gro also mediates the repression produced by dTcf-ΔN, which lacks the Arm-binding region but retains the putative Gro-binding site<sup>19</sup>. Ectopic expression of dTcf-ΔN in wild-type embryos antagonizes Wg signalling. Reduction in maternal Gro levels significantly reduces the ability of dTcf-ΔN to act as a repressor (Fig. 4i, j and Table 1).

Our data indicate that dTcf has a negative as well as a positive



**Figure 4** Gro acts together with dTcf to repress Wg-responsive genes. **a, b**, A reduction in levels of zygotic Gro alone has no effect: *wg*<sup>CX4</sup>; *gro*<sup>BX22</sup>/+ embryos from *gro*<sup>BX22</sup>/+ fathers are indistinguishable from *wg* null mutants in cuticle pattern (**a**, *n* > 100) and *en* expression (**b**, ventral view, stage 13 *wg* mutant). **c**, *wg*<sup>CX4</sup>; *gro*<sup>BX22</sup>/+ embryos from *gro*<sup>BX22</sup>/+ mothers show suppression (203 of 207 embryos). **d**, both *wg*<sup>CX4</sup>; *gro*<sup>BX22</sup>/+ embryos from *gro*<sup>BX22</sup>/+ mothers and *wg*<sup>CX4</sup>; *gro*<sup>E48</sup>/*gro*<sup>BX22</sup> embryos (shown) maintain ventral epidermal *en* expression. **e**, *wg*<sup>CX4</sup>; *gro*<sup>E48</sup>/*gro*<sup>BX22</sup> shows similar rescue of pattern elements compared with *wg* mutants. **f**, *gro*<sup>E48</sup>/*gro*<sup>BX22</sup> embryos show subtle pattern alterations. **g, h**, dTcf

overexpression in *Df(2)DE/wg*<sup>CX4</sup> embryos produces a strong *wg*-like phenotype, even when embryos are zygotically *gro* /+ (**g**, *n* > 100), but not when embryos are maternally *gro* /+ (**h**, 209 of 223 embryos resemble *Df(2)DE/wg*<sup>CX4</sup> embryos). **i, j**, Ubiquitous dTcf-ΔN produces a moderate segment polarity phenotype (**i**), which is suppressed in embryos with *gro* /+ mothers (**j**). **k–o**, Heterozygosity of *gro* substantially suppresses *arm* mutations. **k**, *arm*<sup>XP33</sup>/Y and **l**, *arm*<sup>YD35</sup>/Y embryos show a strong segment polarity phenotype. **m**, *arm*<sup>XP33</sup>/Y; *gro*<sup>BX22</sup>/+, **n**, *arm*<sup>XP33</sup>/Y; *gro*<sup>E48</sup>/+, and **o**, *arm*<sup>YD35</sup>/Y; *gro*<sup>E48</sup>/+, all from *arm* and *gro* heterozygous mothers, show suppression of the strong segment polarity phenotype.

function in transcriptional regulation, and that Gro acts in a repressor complex with dTcf. This dual regulatory role may be conserved in vertebrate Wnt signalling<sup>9,19</sup>. Therefore, we propose that the balance between the activity of Gro and Arm controls cell-fate choice by the Wnt pathway in both vertebrates and invertebrates. □

## Methods

**Fly stocks and crosses.** Cuticle preparations and antibody stainings were performed as described<sup>24</sup>. In Figs 1, 2 and 4, genotypes were assigned by comparing the frequencies of phenotypic classes with expected genotypic frequencies; these data are summarized in Table 1. For *arm*, suppression was documented by ranking embryos in weak to strong phenotypic categories and calculating ratio of embryos in weak categories. *wg<sup>CX4</sup>* is a molecular null allele<sup>25</sup>; *Df(2)DE<sup>18</sup>* is a *wg* hypomorph (A.B., unpublished observations); *arm<sup>XP33</sup>* is a strong hypomorph; *arm<sup>YD35</sup>* is a null allele<sup>15</sup>; both *dTcf* mutations used are molecular null alleles<sup>8</sup>; *gro<sup>E48</sup>* is a putative null point mutation<sup>21</sup>; *gro<sup>BX22</sup>* lacks *gro* and several neighbouring genes in the *Enhancer of split* complex<sup>26</sup>. Gal4 and UAS transgene stocks have been described<sup>8</sup>.

**Mammalian cell culture.** Vector alone (pCDNA3), hTcf-1 or dTcf and Myc-epitope-tagged Gro(1–181) constructs (with a ratio of 10:1) were introduced into COS cells by diethyl aminoethyl-dextran transfections. Cells were prepared for immunohistochemistry using an anti-Myc-antibody.  $2 \times 10^6$  IIAL6 B cells were transfected by electroporation with 1 µg dTcf luciferase reporter plasmid (pTKTOP) or its negative control containing mutated dTcf sites (pTKFOP)<sup>19</sup>. These were co-transfected with 2 µg dTcf expression vector, 0.5 or 5.0 µg Gro expression plasmids and 0.5 µg Arm expression plasmid, balanced to equal plasmid amounts with pCDNA3. Luciferase activity was corrected by chloramphenicol acetyltransferase (CAT) activity<sup>19</sup>. Luciferase and CAT activities were determined as in ref. 8.

Received 15 June; accepted 5 August 1998.

1. Cadigan, K. M. & Nusse, R. Wnt signaling: a common theme in animal development. *Genes Dev.* **11**, 3286–3305 (1997).
2. Bejsovec, A. & Peifer, M. The *wingless/Wnt-1* signaling pathway: new insights into the cellular mechanisms of signal transduction. *Adv. Dev. Biochem.* **4**, 1–45 (1996).
3. Rubinfeld, B. et al. Stabilization of  $\beta$ -catenin by genetic defects in melanoma cell lines. *Science* **275**, 1790–1792 (1997).
4. Morin, P. J. et al. Activation of  $\beta$ -catenin-Tcf signalling in colon cancer by mutations in  $\beta$ -catenin or APC. *Science* **275**, 1787–1790 (1997).
5. Korinek, V. et al. Constitutive transcriptional activation by a  $\beta$ -catenin-Tcf complex in APC(–/–) colon carcinoma. *Science* **275**, 1784–1787 (1997).
6. Brunner, E., Peter, O., Schweizer, L. & Basler, K. *pangolin* encodes a Lef-1 homolog that acts downstream of Armadillo to transduce the Wingless signal. *Nature* **385**, 829–833 (1997).
7. Riese, J. et al. Lef-1, a nuclear factor coordinating signalling inputs from *wingless* and *decapentaplegic*. *Cell* **88**, 777–787 (1997).
8. van de Wetering, M. et al. Armadillo co-activates transcription driven by the product of the *Drosophila* segment polarity gene *dTCF*. *Cell* **88**, 789–799 (1997).
9. Brannon, M., Gomperts, M., Sumoy, L., Moon, R. T. & Kimelman, D. A.  $\beta$ -catenin/XTcf-3 complex binds to the *siamois* promoter to regulate dorsal axis specification in *Xenopus*. *Genes Dev.* **11**, 2359–2370 (1997).
10. DiNardo, S., Kuer, J. M., Theis, J. & O'Farrell, P. H. Development of the embryonic pattern in *D. melanogaster* as revealed by the accumulation of the nuclear *engrailed* protein. *Cell* **43**, 59–69 (1985).
11. DiNardo, S., Sher, E., Heemskerk-Jongens, J., Kassisi, J. A. & O'Farrell, P. H. Two-tiered regulation of spatially patterned *engrailed* gene expression during *Drosophila* embryogenesis. *Nature* **322**, 604–609 (1988).
12. Martinez Arias, A., Baker, N. & Ingham, P. Role of the segment polarity genes in the definition and maintenance of cell states in the *Drosophila* embryo. *Development* **103**, 157–170 (1988).
13. Bejsovec, A. & Martinez-Arias, A. Roles of *wingless* in patterning the larval epidermis of *Drosophila*. *Development* **113**, 471–485 (1991).
14. Heemskerk, J., DiNardo, S., Kostriken, R. & O'Farrell, P. H. Multiple modes of *engrailed* regulation in the progression towards cell fate determination. *Nature* **352**, 404–410 (1991).
15. Peifer, M. & Wieschaus, E. The segment polarity gene *armadillo* encodes a functionally modular protein that is the *Drosophila* homolog of human plakoglobin. *Cell* **63**, 1167–1178 (1990).
16. Von Ohlen, T., Lessing, D., Nusse, R. & Hooper, J. E. Hedgehog signaling regulates transcription through cubitus interruptus, a sequence-specific DNA binding protein. *Proc. Natl. Acad. Sci. USA* **94**, 2404–2409 (1997).
17. Hepker, J., Wang, Q. T., Motzny, C. K., Holmgren, R. & Orenic, T. V. *Drosophila* cubitus interruptus forms a negative feedback loop with patched and regulates expression of Hedgehog target genes. *Development* **124**, 549–558 (1997).
18. Tiong, S. Y. K. & Nash, D. Genetic analysis of the *adenosine3* (*Gart*) region of the second chromosome in *Drosophila melanogaster*. *Genetics* **124**, 889–897 (1990).
19. Roose, J. et al. The *Xenopus* Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. *Nature* **395**, 608–612 (1998).
20. Behrens, J. et al. Functional interaction of  $\beta$ -catenin with the transcription factor Lef-1. *Nature* **382**, 638–642 (1996).
21. Paroush, Z. et al. *groucho* is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with Hairy-related bHLH proteins. *Cell* **79**, 805–815 (1994).
22. de Celis, J. F. & Ruiz-Gomez, M. *groucho* and *hedgehog* regulate *engrailed* expression in the anterior compartment of the *Drosophila* wing. *Development* **121**, 3467–3476 (1995).
23. Noordermeer, J., Johnston, P., Rijsewijk, F., Nusse, R. & Lawrence, P. A. The consequences of ubiquitous expression of the *wingless* gene in the *Drosophila* embryo. *Development* **116**, 711–719 (1992).

24. Bejsovec, A. & Wieschaus, E. Signaling activities of the *Drosophila wingless* gene are separately mutable and appear to be transduced at the cell surface. *Genetics* **139**, 309–320 (1995).
25. Baker, N. E. Molecular cloning of sequences from *wingless*, a segment polarity gene in *Drosophila*: the spatial distribution of a transcript in embryos. *EMBO J.* **6**, 1765–1773 (1987).
26. Schrons, H., Knust, E. & Campos-Ortega, J. A. The *Enhancer of split* complex and adjacent genes in the 96F region of *Drosophila melanogaster* are required for segregation of neural and epidermal progenitor cells. *Genetics* **132**, 481–503 (1992).

**Acknowledgements.** This work was supported by grants from the NIH and the U.S. Army Breast Cancer Research Program (to M.P.), the NSF and the March of Dimes Birth Defects Foundation (to A.B.) and the NWO-GMW (to H.C.).

Correspondence and requests for materials should be addressed to A.B. (e-mail: bejsovec@nrc.nl).

## The *Xenopus* Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors

Jeroen Roose\*, Miranda Molenaar†, Josi Peterson†, Jolanda Hurenkamp\*, Helen Brantjes\*, Petra Moerer\*, Marc van de Wetering\*, Olivier Destree‡ & Hans Clevers\*‡

\* Department of Immunology, University Hospital, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

† Hubrecht Laboratory, Netherlands Institute for Developmental Biology, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

‡ These authors contributed equally to this work

Tcf/Lef transcription factors mediate signalling from Wingless/Wnt proteins by recruiting Armadillo/ $\beta$ -catenin as a transcriptional co-activator<sup>1–7</sup>. However, studies of *Drosophila*, *Xenopus* and *Caenorhabditis elegans* have indicated that Tcf factors may also be transcriptional repressors<sup>6,8–13</sup>. Here we show that Tcf factors physically interact with members of the Groucho family of transcriptional repressors. In transient transfection assays, the *Xenopus* Groucho homologue XGrg-4 inhibited activation of transcription of synthetic Tcf reporter genes. In contrast, the naturally truncated Groucho-family member XGrg-5 enhanced transcriptional activation. Injection of XGrg-4 into *Xenopus* embryos repressed transcription of *Siamois* and *Xnr-3*, endogenous targets of  $\beta$ -catenin-Tcf. Dorsal injection of XGrg-4 had a ventralizing effect on *Xenopus* embryos. Secondary-axis formation induced by a dominant-positive Armadillo-Tcf fusion protein was inhibited by XGrg-4 and enhanced by XGrg-5. These data indicate that expression of Tcf target genes is regulated by a balance between Armadillo and Groucho.

In our yeast two-hybrid screen for proteins interacting with human (h) TCF-1, which led to the identification of  $\beta$ -catenin<sup>2</sup>, roughly 60 out of 300 clones encoded the murine (Gro)-related gene *Grg-5* (ref. 14). We confirmed independently that TCF-1 and Grg-5 interact in a binding assay using a recombinant maltose-binding protein (MBP)–Grg5 fusion protein and *in vitro*-translated hTCF-1 (Fig. 1a).

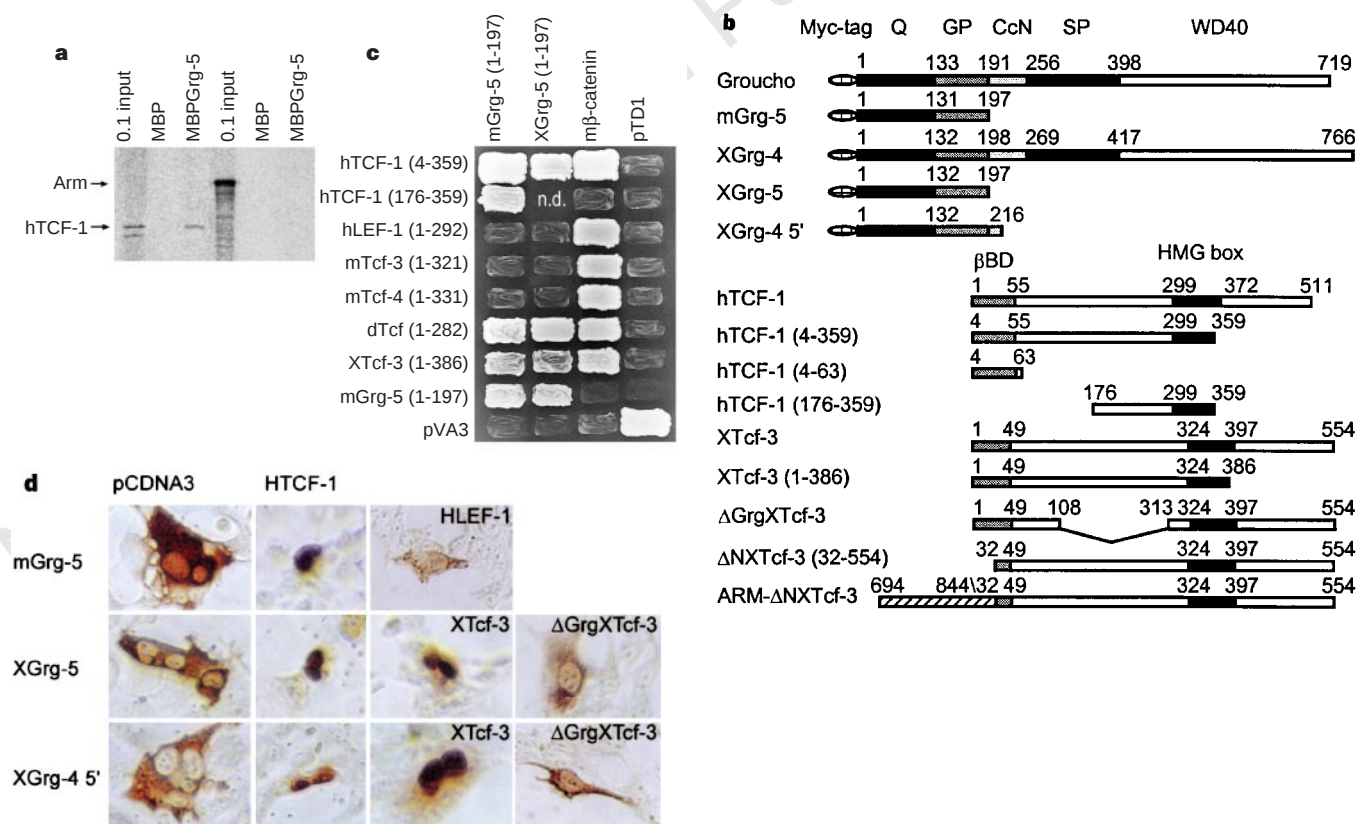
*Groucho* (*Gro*) is a broadly expressed *Drosophila* corepressor, and may be involved in segmentation, sex determination and neurogenesis<sup>15–18</sup>. Hairy and Enhancer of Split-like (HES) helix–loop–helix factors interact with the non-DNA-binding Gro protein to repress transcription of their target genes<sup>16,19</sup>. In mammals, multiple homologues with a similar overall domain structure have been identified. These are termed *TLE 1–4* in man, and *mGrg-1*, *-3* and *-4* in mouse (Fig. 1b). *Grg-5* encodes only the two amino-terminal domains of these proteins (Fig. 1b). In a search for maternally expressed Gro proteins in *Xenopus*, we cloned XGrg-5 (or XAES<sup>20</sup>) and a *Grg-4* orthologue (Fig. 1b). Both were abundantly and ubiquitously expressed in oocytes and in embryos undergoing the pre-midblastula transition (results not shown).

In yeast two-hybrid assays, we found that mGrg-5 and XGrg-5 interacted with *Drosophila* (d) Tcf, XTcf-3 and, as reported previously<sup>21</sup>, with mGrg-5 itself (Fig. 1c), but not with hLEF-1, mTcf-3 and mTcf-4. Deletion analysis defined a minimal region in hTCF-1 (amino acids 176–359) that was capable of binding to Grg-5; this domain was separable from the Armadillo (Arm)-interaction domain (amino acids 4–63; ref. 22). No conclusive data were obtained in yeast for the interactions between Tcfs and the 'long' Gro homologues, which we attribute to transcriptional repression of the selection gene (histidine). To circumvent this problem, tagged N-terminal fragments of Gro proteins (collinear with full-length Grg-5; Fig. 1b) were expressed in COS cells. Although the full-length proteins were nuclear (not shown), these N-terminal fragments localized to the cytoplasm. Co-transfection of such truncated complementary DNA clones with various Tcf expression constructs allowed us to visualize interaction between Tcfs and the long Gro homologues by nuclear translocation. In this assay, hTCF-1 interacted with mGrg-5, XGrg-4 and XGrg-5, whereas XTcf-3 interacted with XGrg-4 and XGrg-5. Removal of the Grg-interaction domain from XTcf-3 (Fig. 1b;  $\Delta$ Grg-XTcf-3) abrogated the interaction of XTcf-3 with Gro homologues and nuclear translocation of the complexes (Fig. 1d).

In transient transfections using a previously established  $\beta$ -catenin–Tcf reporter gene assay<sup>5</sup>, we found that XGrg-4 repressed the

activation of transcription by Arm and XTcf-3 (Arm–XTcf3 complexes) (Fig. 2a). The repression was specific, as we did not observe effects on the mutant Tcf reporter gene (Fig. 2a), nor on the co-transfected control chloramphenicol acetyltransferase (CAT) vector. In contrast, XGrg-5, which lacks the carboxy-terminal WD40 repeats of the longer Grg proteins<sup>14</sup>, enhanced the transcriptional activity of suboptimal amounts of Arm–XTcf-3 complexes (Fig. 2b). mGrg-5 had no intrinsic transactivation properties when fused to a Gal4 DNA-binding domain (not shown). The enhancement of transcription by XGrg-5 could probably be attributed to its interference with the repressive effects of endogenous Gro proteins. We note that each line in a large and diverse cell panel expressed multiple *Grg/TLE* genes (H.B., J.R. and H.C., unpublished observations); the B-cell line used in our reporter assay expressed moderate levels of both *Grg-1* and *Grg-4* (results not shown). A deletion mutant of XTcf-3 that lacked the Grg-interaction domain was a tenfold more potent transcriptional activator than its wild-type counterpart (Fig. 2c), confirming the activity of endogenous co-repressors of Tcf factors in our assay. As expected, this  $\Delta$ Grg-XTcf-3 mutant was not subject to repression by XGrg-4 or anti-repression by XGrg-5 (Fig. 2d).

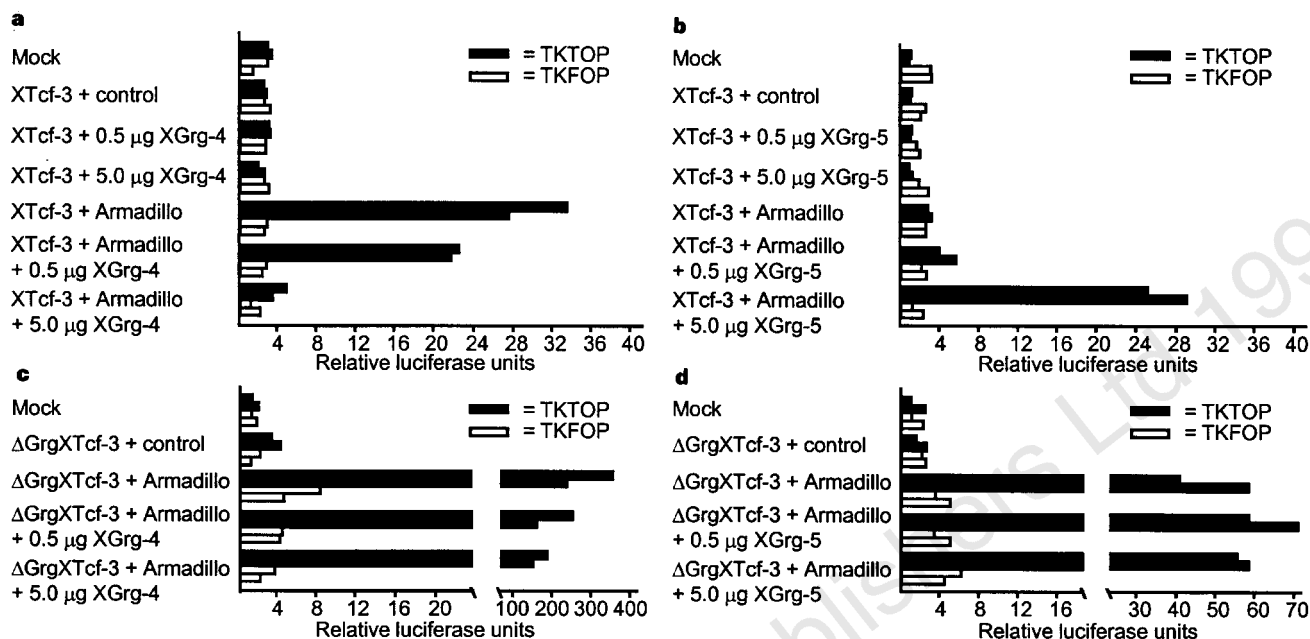
We then determined whether the Gro–Tcf interaction was involved in the *in vivo* regulation of  $\beta$ -catenin–Tcf target genes in *Xenopus* embryos. Tcf-binding sites in the *Siamois* promoter have



**Figure 1** Physical interaction between Groucho (Gro)-related proteins and Tcf. **a**, *In vitro* binding assay for hTCF-1 and Grg-5. *In vitro*-transcribed and translated hTCF-1 binds to an MBP-Grg-5 fusion protein (lane 3), but not to control MBP protein (lane 2). *In vitro*-translated Armadillo does not bind to these MBP proteins (lane 5, 6). Input hTCF-1 protein is run in lane 1, and input Armadillo protein in lane 4. **b**, Domains of Gro and Tcf constructs. Like Gro, XGrg-4 contains five distinct domains: mGrg-5 and XGrg-5 contain only the Q and GP domains<sup>14</sup>. hTCF-1 and XTcf-3 contain a centrally located DNA-binding HMG box and the N-terminal  $\beta$ -catenin-binding domain ( $\beta$ BD).  $\Delta$ Grg-XTcf-3 was constructed by deleting amino acids 109–312 of XTcf-3. A dominant-positive version of XTcf-3 (bottom) was created in which its N terminus is replaced by the C-terminal transactivation

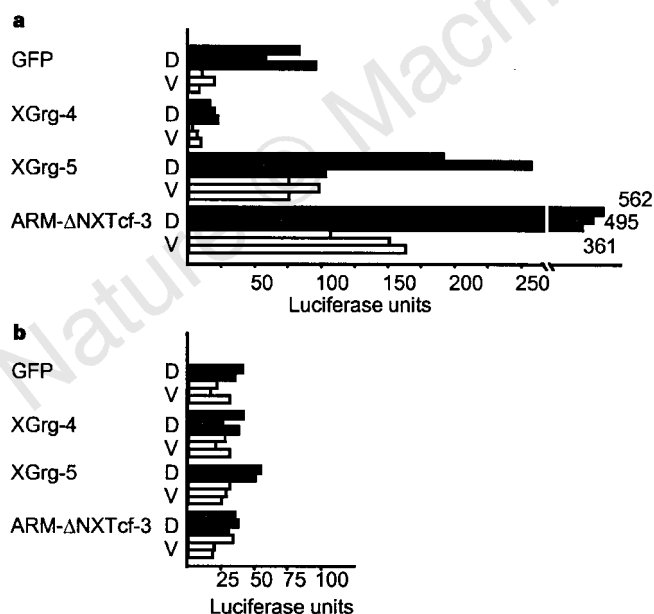
domain of Arm (amino acids 694–844). **c**, Two-hybrid assay for the interaction of Tcfs, Grgs and  $\beta$ -catenin. All tested Tcf-family members bind to  $\beta$ -catenin. In contrast, only hTCF-1, dTcf, and XTcf-3 interact with Grg-5. The Grg5-interaction domain of Tcf proteins (amino acids 176–359) is separable from the domain that interacts with  $\beta$ -catenin (amino acids 4–63 (ref. 22)). **d**, Tcf transports Gro to the nucleus. Tagged, truncated Gro proteins localize to the cytoplasm of COS cells. Introduction of hTCF-1 results in nuclear translocation of mGrg-5, XGrg-5 and XGrg-4 5'. In the same way, XTcf-3 interacts with XGrg-5 and XGrg-4 5'.  $\Delta$ Grg-XTcf-3, however, did not cause nuclear translocation of XGrg-5 and XGrg-4 5'. Introduction of hLEF-1 does not result in nuclear translocation of mGrg-5.



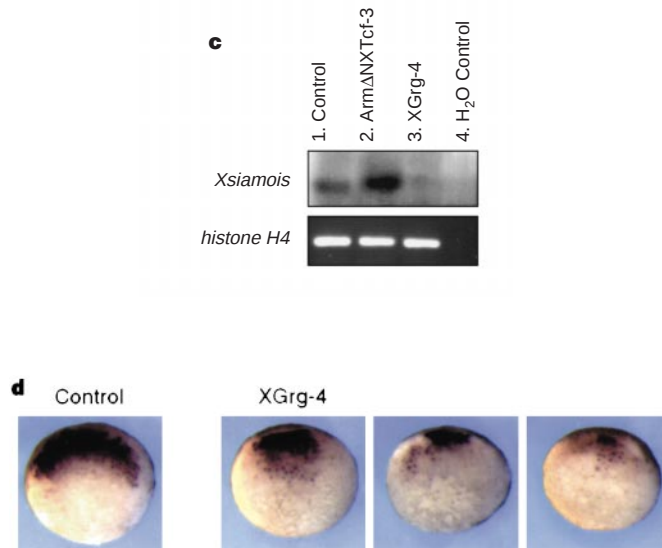


**Figure 2** Gro represses Arm-Tcf-driven transactivation of a Tcf reporter. **a**, IIA1.6 B cells were transfected with optimal amounts of the indicated plasmids. XGrg-4 represses transactivation by Arm-XTcf-3. Tenfold transactivation is induced by 5 µg Arm plasmid. **b**, XGrg-5 enhances transcription in the presence of a suboptimal amount of Arm (0.5 µg). **c**, 5 µg Arm plasmid induces a 100-fold transactivation, with ΔGrg-XTcf-3 as an effector protein. The transactivation

cannot be repressed by introduction of XGrg-4. **d**, XGrg-5 does not enhance transcription induced by ΔGrg-XTcf-3 and Arm (0.5 µg). In all figures, both values of duplicate transfections are given (corrected for transfection efficiency based on an internal CAT control). TKTOP, Tcf Optimal reporter; TKFOP, Tcf Far-from-Optimal reporter (see Methods).



**Figure 3** XGrg-4 represses *Siamese* promoter activity *in vivo*. **a**, **b**, For *Siamese* reporter assays two different constructs, S01234 and S24 (ref. 11), were injected in the equatorial region of the two dorsal (D) or the two ventral (V) blastomeres at the four-cell stage. S01234 is a luciferase reporter construct consisting of 0.8 kilobases of the wild-type *Siamese* promoter, containing three Tcf consensus sites that are β-catenin-responsive; these three sites are mutated in S24 (ref. 11). **a**, 265 pg S01234 were injected together with equimolar amounts of capped synthetic mRNA, encoding green fluorescent protein (GFP) (750 pg per embryo), ArmΔNXTcf-3 (500 pg), XGrg-4 (3,000 pg) or XGrg-5 (3,000 pg). Luciferase activities from two to three pools of five stage-10 embryos were determined for each combination. XGrg-4 represses dorsal *Siamese* promoter activation, whereas XGrg-5 enhances the activity. **b**, No effects are seen when the different RNAs are injected together with S24. **c**, RT-PCR analysis of *Siamese* expression.



Embryos were injected at the four-cell stage in the equatorial region of each blastomere with either 500 pg ArmΔNXTcf-3 (ventral) or 3,000 pg XGrg-4 (dorsal). Twenty embryos were pooled for each data point. Dorsal injection of XGrg-4 RNA results in a strong reduction of *Siamese* expression (lane 3), whereas *Siamese* expression is enhanced on injection of ArmΔNXTcf-3 (lane 2). The amount of cDNA per sample was standardized for histone H4 expression (bottom). Nearly identical results were obtained in each of three independent experiments. Control embryos raised until stage 42 showed typical phenotypes (Fig. 4; ref. 2). **d**, *In situ* hybridization with *Xnr-3* for stage-9 embryos injected dorso-equatorially with 3,000 pg XGrg-4 RNA at the four-cell stage. Left to right: non-injected embryo, in which expression of *Xnr-3* is seen at the future dorsal side, and three embryos injected with increasing amounts of XGrg-4, showing increasingly reduced levels of *Xnr-3* RNA.

| Position   | n   | DAI | Percent | Phenotypes |         |  |
|------------|-----|-----|---------|------------|---------|--|
| Animal     | 216 | 5   | 32      |            | control |  |
|            |     | 4   | 60      |            |         |  |
|            |     | 3   | 1       |            |         |  |
| Equatorial | 272 | 5   | 18      |            | XGrg-4  |  |
|            |     | 4   | 37      |            |         |  |
|            |     | 3   | 29      |            |         |  |
|            |     | 2   | 1       |            |         |  |
|            |     | 1   | 1       |            |         |  |
| Vegetal    | 201 | 5   | 38      |            |         |  |
|            |     | 4   | 50      |            |         |  |
|            |     | 3   | 1       |            |         |  |

**Figure 4** Dorsal injection of *XGrg-4* suppresses endogenous axis formation. Embryos were injected, at the four-cell stage, in each dorsal blastomere with 1,500 pg *XGrg-4* at the position indicated. The embryos were scored at stages 33–40 according to the standard dorso-anterior index scale (DAI<sup>30</sup>). A normal embryo is assigned DAI 5, whereas embryos lacking dorso-anterior structures are assigned DAI 0. A typical range of phenotypes found in one experiment is shown at the right. No effects were observed upon injection of *XGrg-5* or of control *GFP* or  $\beta$ -galactosidase mRNA.

been proposed to mediate ventral repression of this  $\beta$ -catenin-regulated gene in *Xenopus* embryos<sup>10,11</sup>. When we injected this promoter into *Xenopus* embryos, we obtained results that were nearly identical to those obtained when expressing the synthetic reporters in mammalian cells: *XGrg-4* repressed promoter activity, whereas *XGrg-5* enhanced the activity (Fig. 3a). These effects depended on the presence of the three Tcf-binding sites in the promoter (Fig. 3b). We then studied the consequence of injection of *XGrg-4* on the expression of the endogenous *Siamois* gene. Because it is expressed at low levels, we used the semiquantitative polymerase chain reaction (PCR) to amplify *Siamois* RNA purified from pooled, *XGrg-4*-injected embryos<sup>23</sup>. Injection of *XGrg-4* at the four-cell stage led to a strong reduction of transcription of the endogenous *Siamois* gene at stage 10, whereas injection of dominant-positive XTcf-3 (see below) enhanced transcription of *Siamois* (Fig. 3c).

To test the effect of *XGrg-4* on another direct target of XTcf-3, *Xnr-3* (ref. 24), we performed whole-mount *in situ* hybridization<sup>25</sup> on stage 9 embryos. As expected, dorsal injections of *XGrg-4* RNA at the four-cell stage markedly reduced the *Xnr-3* signal (Fig. 3d).

We have shown previously that XTcf-3 in *Xenopus* embryos mediates the axis-inducing  $\beta$ -catenin signal<sup>2</sup>. As *XGrg-4* and *XGrg-5* both interact with XTcf-3 and are expressed maternally, we analysed the effect of dorsal injections of these two proteins on axis formation. Injection of *XGrg-4* inhibited formation of the endogenous axis, resulting in ventralization of the embryos, whereas *XGrg-5* had no effect. Effects were strongest after injection in the equatorial region of the two dorsal blastomeres of four-cell-stage embryos. Phenotypes ranged from complete lack of a head in combination with a shortened tail, to microcephaly and cyclopia (Fig. 4).

We also tested the effects of Gro on axis duplication. To avoid perturbations of pools of endogenous Wnt cascade components, we designed a dominant-positive version of XTcf-3 in which the N terminus is replaced by the C-terminal transactivation domain of Arm<sup>5</sup> (Fig. 1b). Although ventrally injected XTcf-3 is essentially inert and  $\Delta$ N-XTcf-3 is a potent dominant-negative mutant<sup>2</sup>, the chimaeric protein potentially induced secondary axes (Table 1). This showed that recruitment of the transactivating C terminus of Arm to Tcf sites constitutes the primary nuclear event upon signalling. Injection of the chimaeric protein with *XGrg-4* inhibited this activity, but injection of the chimaera with *XGrg-5* potentiated the activity (Table 1). This result concurred with the repressive effects of *XGrg-4* and the enhancing effects of *XGrg-5* on transcription in mammalian cells. Ventral injections of *XGrg-5* RNA alone had

**Table 1** *XGrg-4* suppresses, *XGrg-5* potentiates axis duplication by Arm

| RNA injected (pg)                               | Incomplete secondary axis (%) | Secondary axis incl. eye and cement gland (%) | Total number of embryos injected |
|-------------------------------------------------|-------------------------------|-----------------------------------------------|----------------------------------|
| 500 Arm- $\Delta$ NXTcf-3 + 1,500 <i>XGrg-4</i> | 17                            | 0                                             | 334                              |
| 500 Arm- $\Delta$ NXTcf-3 + 375 GFP             | 40                            | 18                                            | 357                              |
| 25 Arm- $\Delta$ NXTcf-3 + 1,000 <i>XGrg-5</i>  | 25                            | 4                                             | 408                              |
| 25 Arm- $\Delta$ NXTcf-3 + 1,000 GFP            | 9                             | 2                                             | 384                              |
| 100 <i>XNoggin</i> + 1,500 <i>XGrg-4</i>        | 46                            | 6                                             | 282                              |
| 100 <i>XNoggin</i> + 375 GFP                    | 46                            | 8                                             | 269                              |

Embryos were injected at the four-cell stage in the equatorial region of one ventral blastomere and screened for secondary axis induction at stages 25–30. As a negative control, equimolar amounts of green fluorescent protein (GFP) mRNA were injected.

no effect on axis duplication (results not shown), indicating that derepression alone is not sufficient for the biological effect. Axis duplication induced by injection of *noggin* messenger RNA<sup>26</sup> could not be blocked by *XGrg-4* (Table 1).

We propose that the transcription of Tcf target genes is the result of a balance between the constitutive, repressive effects mediated by Gro (possibly counteracted by Grg5-like proteins) and the activating effects of Arm. In non-signalling cells, the repressive activities will dominate. Following Wnt activation,  $\beta$ -catenin will associate with Tcfs and will activate transcription of genes such as *Xnr-3* and *Siamois* in *Xenopus*, or *Engrailed* and *Ubx* in flies. The active repression by Gro secures tight control over the Wntless/Wnt-driven developmental decisions. The dual activities of Tcf factors may explain the puzzling observation that mutation of the *C. elegans* Tcf homologue *Pop-1* has opposite effects on E versus MS-cell specification to those resulting from mutation of Wnt and Arm<sup>12</sup>. *Pop-1* probably functions mainly as a repressor of target gene transcription. Our results predict a role for Gro in *Pop-1*-controlled cell-fate decisions.

Constitutive activation of Tcf-mediated transcription occurs in melanoma and colon carcinoma after loss of APC or gain-of-function mutations in  $\beta$ -catenin<sup>27–29</sup>. As Gro proteins repress Tcf activity, it will be interesting to study the status of *TLE* genes in human cancers in which Tcf transcription is inappropriately activated. □

## Methods

**Two-hybrid experiments** were performed as described<sup>2</sup>. Plasmid pVA3 encodes a murine p53–Gal4 binding domain hybrid in pGBT9 (Clontech). Preys mGrg-5, *XGrg-5*, and  $\beta$ -catenin were inserted in frame with the Gal4 activation domain in pGADGH (Clontech) or pGADRX (Stratagene). pTD1 encodes SV40 large T antigen in pGAD3F (Clontech). Baits and preys were transformed into the *Saccharomyces cerevisiae* strain HF7C (Clontech).

**In vitro binding assays.** Radiolabelled hTcf-1 and Armadillo were produced in the PROTEINscript kit (Ambion). Translated products were diluted in 0.5 ml ELB buffer (150 mM NaCl, 50 mM HEPES, pH 7.5, 5 mM EDTA, 0.1% NP40, 10 mM  $\beta$ -glycerophosphate, 5 mM NaF, 1 mM PMSF, 10 mg ml<sup>−1</sup> trypsin inhibitor, 20 U ml<sup>−1</sup> aprotinin and 1 mM sodium orthovanadate). Amylose-sepharose beads loaded with MBP–Grg5 fusion protein or control MBP were incubated at 4°C for 2 h. Washed beads were analysed by gel electrophoresis and autoradiography.

**Cloning of *XGrg-5* and *XGrg-4*.** A *Xenopus* brain complementary DNA library in  $\lambda$ gt11 (ref. 2) was screened at low stringency with murine *Grg-5* cDNA probes according to standard procedures. *XGrg-5* was previously described as XAES<sup>20</sup> (GenBank accession number U18776). The accession number of *XGrg-4* is AJ224945.

**COS cell transfections.** COS cells were transfected with Tcf and *XGrg* constructs in a ratio of 10:1 using standard diethyl aminoethyl (DEAE)-dextran transfection. After 24 h, cells were methanol-fixed and stained using an anti-Myc-tag antibody.

**Luciferase assays and CAT assays.**  $2 \times 10^6$  IIAI.6 B cells were electroporated with a luciferase reporter plasmid containing three optimal Tcf sites upstream of the minimal HSV-TK promoter (1  $\mu$ g pTKTOP) or its negative control containing mutated Tcf sites (pTKFOP), the internal transfection control (0.5  $\mu$ g pSV40CAT), 20  $\mu$ g Tcf expression vectors, and 0.5 or 5.0  $\mu$ g Gro expression plasmids. For XGrg-5 experiments, 0.5  $\mu$ g Arm was used; for XGrg-4 experiments, 5.0  $\mu$ g pCDNA, was used as a stuffer when appropriate. cDNAs encoding tagged versions of XTcf-3,  $\Delta$ Grg-XTcf-3, XGrg-4 and XGrg-5 were inserted into pCDNA3. Luciferase and CAT activities were separately determined 24 h after transfection as described<sup>5</sup>.

**Reverse transcription (RT)-PCR analysis.** Total RNA for detection of endogenous *Siamois* mRNA by RT-PCR was isolated from stage 10 embryos<sup>2</sup>. Oligo-d(T)-primed cDNA from total RNA was prepared using standard techniques. cDNA was quantified by PCR analysis for histone H4 and products were compared after 12, 14, 16, 18 and 20 cycles. The product after 20 cycles is shown in Fig. 3c. RT-PCR was carried out as described<sup>23</sup>.

Received 11 June; accepted 30 July 1998.

- Behrens, J. *et al.* Functional interaction of  $\beta$ -catenin with the transcription factor Lef-1. *Nature* **382**, 638–642 (1996).
- Molenaar, M. *et al.* XTcf-3 transcription factor mediates beta-catenin-induced axis formation in *Xenopus* embryos. *Cell* **86**, 391–399 (1996).
- Huber, O. *et al.* Nuclear localization of beta-catenin by interaction with transcription factor Lef-1. *Mech. Dev.* **59**, 3–10 (1996).
- Brunner, E., Peter, O., Schweizer, L. & Basler, K. *pangolin* encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in *Drosophila*. *Nature* **385**, 829–833 (1997).
- van de Wetering, M. *et al.* Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene dTCE. *Cell* **88**, 789–799 (1997).
- Riese, J. *et al.* Lef-1, a nuclear factor coordinating signaling inputs from wingless and decapentaplegic. *Cell* **88**, 777–787 (1997).
- Cadigan, K. M. & Nusse, R. Wnt signaling: a common theme in animal development. *Genes Dev.* **11**, 3286–3305 (1997).
- Thorpe, C. J., Schlesinger, A., Carter, J. C. & Bowerman, B. Wnt signaling polarizes an early *C. elegans* blastomere to distinguish endoderm from mesoderm. *Cell* **90**, 695–705 (1997).
- Rochelleau, C. E. *et al.* Wnt signaling and an APC-related gene specify endoderm in early *C. elegans* embryos. *Cell* **90**, 707–716 (1997).
- Lemaire, P. *et al.* Expression cloning of *Siamois*, a *Xenopus* homeobox gene expressed in dorsal-vegetal cells of blastulae and able to induce a complete secondary axis. *Cell* **81**, 85–94 (1995).
- Brannon, M., Gomperts, M., Sumoy, L., Moon, R. T. & Kimelman, D. A beta-catenin/XTcf-3 complex binds to the *siamois* promoter to regulate dorsal axis specification in *Xenopus*. *Genes Dev.* **11**, 2359–2370 (1997).
- Han, M. Gut reaction to Wnt signaling in worms. *Cell* **90**, 581–584 (1997).
- Bienz, M. TCF: transcriptional activator or repressor? *Curr. Opin. Cell Biol.* **10**, 366–372 (1998).
- Mallo, M., Franco del Amo, F. & Gridley, T. Cloning and developmental expression of Grg, a mouse gene related to the groucho transcript of the *Drosophila* Enhancer of split complex. *Mech. Dev.* **42**, 67–76 (1993).
- Hartley, D. A., Preiss, A. & Artavanis Tsakonas, S. A deduced gene product from the *Drosophila* neurogenic locus, enhancer of split, shows homology to mammalian G-protein beta subunit. *Cell* **55**, 785–795 (1988).
- Paroush, Z. *et al.* Groucho is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* **79**, 805–815 (1994).
- Parkhurst, S. M. Groucho: making its Marx as a transcriptional co-repressor. *Trends Genet.* **14**, 130–132 (1998).
- Fisher, A. L. & Caudy, M. Groucho proteins: transcriptional corepressors for specific subsets of DNA-binding transcription factors in vertebrates and invertebrates. *Genes Dev.* **12**, 1931–1940 (1998).
- Jimenez, G., Paroush, Z. & Ish Horowitz, D. Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. *Genes Dev.* **11**, 3072–3082 (1997).
- Choudhury, B. K., Kim, J., Kung, H. F. & Li, S. S. Cloning and developmental expression of *Xenopus* cDNAs encoding the enhancer of split groucho and related proteins. *Gene* **195**, 41–48 (1997).
- Pinto, M. & Lobe, C. G. Products of the grg (Groucho-related gene) family can dimerize through the amino-terminal Q domain. *J. Biol. Chem.* **271**, 33026–33031 (1996).
- Korinek, V. *et al.* Two members of the Tcf family implicated in Wnt/ $\beta$ -catenin signaling during embryogenesis in the mouse. *Mol. Cell. Biol.* **18**, 1248–1256 (1998).
- Laurent, M. N., Blitz, I. L., Hashimoto, C., Rothbächer, U. & Cho, K. W. Y. The *Xenopus* homeobox gene *Twin* mediates Wnt induction of *Gooseoid* in establishment of Spemann's organizer. *Development* **124**, 4905–4916 (1997).
- McKendry, R., Hsu, S. C., Harland, R. M. & Grosschedl, R. Lef-1/TCF proteins mediate wnt-inducible transcription from the *Xenopus* nodal-related 3 promoter. *Dev. Biol.* **192**, 420–431 (1997).
- Harland, R. M. *et al.* In situ hybridization: an improved whole mount method for *Xenopus laevis*. *Methods Enzymol.* **36**, 685–697 (1991).
- Smith, W. C. & Harland, R. M. Expression cloning of noggin, a new dorsalizing factor localized to the Spemann organizer in *Xenopus* embryos. *Cell* **70**, 829–840 (1992).
- Korinek, V. *et al.* Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science* **275**, 1784–1787 (1997).
- Morin, P. J. *et al.* Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science* **275**, 1787–1790 (1997).
- Rubinfeld, B. *et al.* Stabilization of beta-catenin by genetic defects in melanoma cell lines. *Science* **275**, 1790–1792 (1997).
- Kao, K. & Elinson, R. P. The entire mesodermal mantle behaves as Spemann's organizer in dorso-anterior enhanced *Xenopus laevis* embryos. *Dev. Biol.* **127**, 64–77 (1988).

**Acknowledgements.** We thank M. Peifer and A. Bejsovec for sharing unpublished information; C. Lobe and K. Koop for Grg cDNAs; S. Stifani for TLE cDNAs and helpful comments; W. C. Smith for *Xnr3* and *Xnogg* cDNA; D. Kimelman for *siamois* reporter constructs; and R. Medema for biochemical advice.

Correspondence and requests for materials should be addressed to H.C. (e-mail: HClivers@LAB.AZU.NL).

## Androstane metabolites bind to and deactivate the nuclear receptor CAR- $\beta$

Barry M. Forman<sup>\*†</sup>, Iphigenia Tzamelis<sup>‡</sup>, Hueng-Sik Choi<sup>§</sup>, Jasmine Chen<sup>\*†</sup>, Devendranath Simha<sup>||</sup>, Wongi Seol<sup>||</sup>, Ronald M. Evans<sup>†</sup> & David D. Moore<sup>‡</sup>

<sup>\*</sup> The City of Hope National Medical Center, 1500 East Duarte Road, Duarte, California 91010, USA

<sup>†</sup> The Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 Torrey Pines Road, La Jolla, California 92037, USA

<sup>‡</sup> Department of Cell Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77401, USA

<sup>§</sup> Hormone Research Center, Chonnam National University, Kwangju 500-757, Republic of Korea

<sup>||</sup> Department of Molecular Biology, Wellman 913, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

<sup>¶</sup> Division of Neoplastic Disease Mechanisms, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA

The orphan receptor CAR- $\beta$  (ref. 1) binds DNA as a heterodimer with the retinoid-X receptor and activates gene transcription in a constitutive manner. Here we show that, in contrast to the classical nuclear receptors, the constitutive activity of CAR- $\beta$  results from a ligand-independent recruitment of transcriptional co-activators. While searching for potential ligands of CAR- $\beta$ , we found that the steroids androstanol and androstenol inhibit the constitutive activity of CAR- $\beta$ . This effect is stereospecific: only 3 $\alpha$ -hydroxy, 5 $\alpha$ -reduced androstanes are active. These androstanes do not interfere with heterodimerization or DNA binding of CAR- $\beta$ ; instead, they promote co-activator release from the ligand-binding domain. These androstane ligands are examples of naturally occurring inverse agonists<sup>2,3</sup> that reverse transcriptional activation by nuclear receptors. CAR- $\beta$  (constitutive androstane receptor- $\beta$ ), therefore, defines an unanticipated steroidal signaling pathway that functions in a manner opposite to that of the conventional nuclear receptor pathways.

Unlike classical nuclear hormone receptors which are activated by their cognate ligands, mouse (m) CAR- $\beta$  (ref. 1) is transcriptionally active in the absence of exogenous hormone (Fig. 1a). As the ligand-binding domain (LBD) of CAR- $\beta$  contains sequence determinants characteristic of classical nuclear receptors, we determined whether CAR- $\beta$  could respond to exogenous signalling molecules. Surprisingly, the constitutive activity of CAR- $\beta$  was completely inhibited by the mammalian pheromone 5 $\alpha$ -androst-16-en-3 $\alpha$ -ol (ref. 4) (androstenol, 5  $\mu$ M, Fig. 1a, b) and by 5 $\alpha$ -androst-3 $\alpha$ -ol (androstanol, 5  $\mu$ M, Fig. 1a, b). Both compounds exhibit half-maximal inhibition at concentrations of about 400 nM (Fig. 1c). Inhibition was specific for CAR- $\beta$ , as no inhibition was observed with other receptors (data not shown).

Further studies showed that inhibition is stereospecific for 5 $\alpha$ -reduced compounds with a 3 $\alpha$ -hydroxy group. A compound lacking the 3-hydroxy moiety (5 $\alpha$ -androstane) was entirely inactive, whereas compounds with a 3 $\beta$ -hydroxy group (5 $\alpha$ -androst-3 $\beta$ -ol, 5 $\alpha$ -androst-16-en-3 $\beta$ -ol) or a 3-keto group (androstenone, 5 $\alpha$ -androst-16-ene-3-one) had EC<sub>50</sub> values (effector concentrations for a half-maximal response) of greater than 10  $\mu$ M (Fig. 1c; data not shown). Similarly, potency is reduced by more than tenfold when a 5 $\beta$ -reduced compound (5 $\beta$ -androst-3 $\alpha$ -ol) is used. This pattern of response indicates that *in vivo* production of a CAR- $\beta$  inhibitor may require the activity of steroid 5 $\alpha$ -reductase.

The fact that both androstanol and androstenol reverse transcriptional activation by CAR- $\beta$  indicated that other modifications might be tolerated at the 16 and 17 positions. However, many



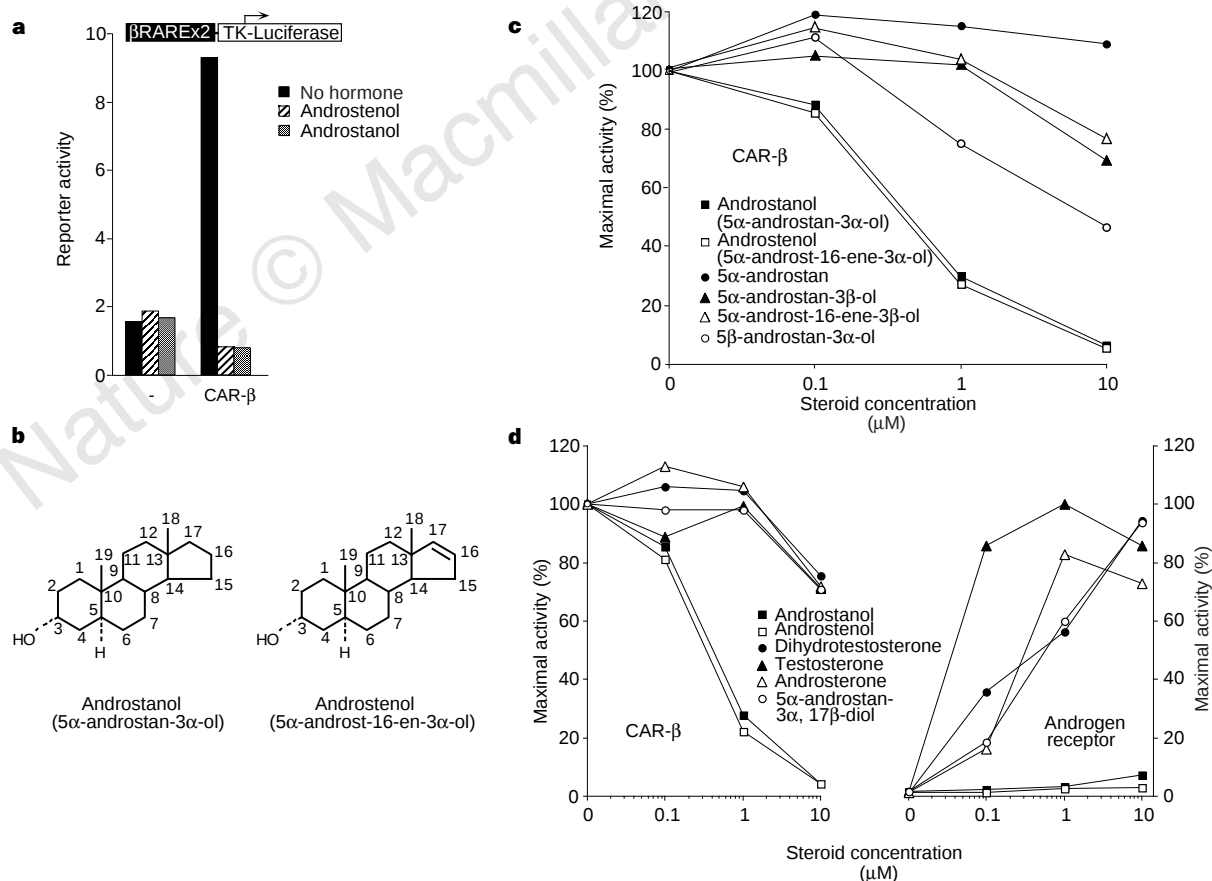
other compounds showed little or no activity, including androsterone (5 $\alpha$ -androstan-3 $\alpha$ -ol-17-one) and 5 $\alpha$ -androstan-3 $\alpha$ ,17 $\beta$ -diol (Fig. 1d), dehydroepiandrosterone (DHEA), 5 $\alpha$ -cholestan-3 $\alpha$ -ol and pregnenolone (data not shown). True androgens such as testosterone and dihydrotestosterone do not inhibit CAR- $\beta$  activity, and neither androstenol nor androstanol is able to activate the androgen receptor (Fig. 1d). Although androstenol is a mammalian pheromone, the CAR- $\beta$  structure-activity profile differs from that of the pheromonal response to these compounds<sup>5</sup>. These results indicate that CAR- $\beta$  is a constitutive androstane receptor that is responsive to a distinct class of naturally occurring androstane metabolites.

Classical nuclear hormone receptors contain a modular LBD that can confer ligand-responsiveness to heterologous DNA-binding domains (DBDs). Thus, we determined whether the inhibitory effects of androstanes are inherent to the CAR- $\beta$  LBD. A chimaeric receptor containing the LBD of CAR- $\beta$  fused to the yeast GAL4 DBD exhibits constitutive activity that is blocked by androstanol (Fig. 2a). Full constitutive activity required co-expression of the LBD of retinoid-X receptor- $\alpha$  (RXR- $\alpha$ ) (Fig. 2a), the heterodimeric partner common to CAR and other nuclear receptors. This indicates that, unlike the classical ligand-dependent nuclear receptors, a complex containing the LBDs of CAR- $\beta$  and RXR- $\alpha$  confers constitutive activity on a heterologous DBD. One explanation for such activity might be the presence of endogenous activators for the different orphan receptors in mammalian cells<sup>6</sup>. Thus, we asked whether the CAR $\beta$ -RXR $\alpha$  complex also exhibits constitutive activ-

ity in yeast, which lack endogenous steroid receptor signalling pathways. As in mammalian cells, the CAR- $\beta$  LBD/RXR- $\alpha$  LBD heterodimer exhibited both constitutive activity and androstanol-dependent repression in yeast (Fig. 2b). The inhibitory effect of the androstanes is, therefore, inherent to the CAR $\beta$ -RXR $\alpha$  heterodimer and cannot be ascribed to any other mammalian-specific activities.

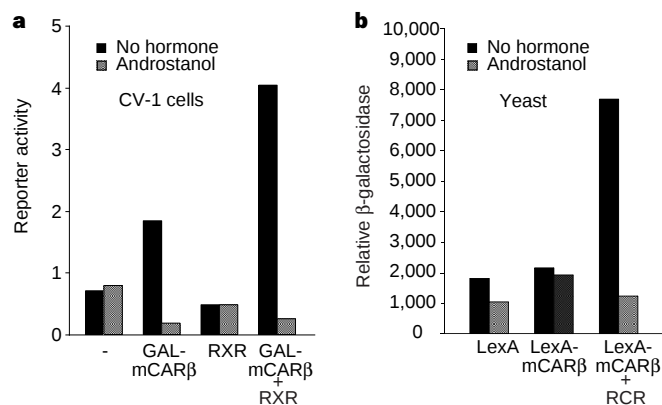
The results obtained using chimaeras containing heterologous DBDs indicate that repression by androstanol may not reflect an inhibition of DNA binding. Consistent with this, androstanol had no effect on DNA binding by CAR $\beta$ -RXR $\alpha$  heterodimers *in vitro* and CAR $\beta$ -RXR $\alpha$  heterodimers from androstanol-treated cells retained the ability to bind DNA (data not shown). Similarly, two-hybrid studies in yeast and mammalian cells indicate that the CAR $\beta$ -RXR $\alpha$  interaction is not disrupted by androstanol (data not shown). These data are consistent with the theory that CAR- $\beta$  is truly a ligand-independent transcriptional activator.

Activation of classical nuclear receptors occurs through a ligand-induced conformational change that promotes recruitment of transcriptional co-activators such as SRC-1 (ref. 7). In principle, the activities of CAR- $\beta$  might be explained by the reverse phenomenon, that is, ligand-independent recruitment of a co-activator followed by androstane-induced co-activator release. Consistent with this proposal, two-hybrid assays showed that co-expression of VP16-CAR $\beta$  (where VP16 is the transactivation domain of the herpesvirus VP16 protein) and GAL4-SRC1 in mammalian cells resulted in a functional interaction, which was reversed by andros-

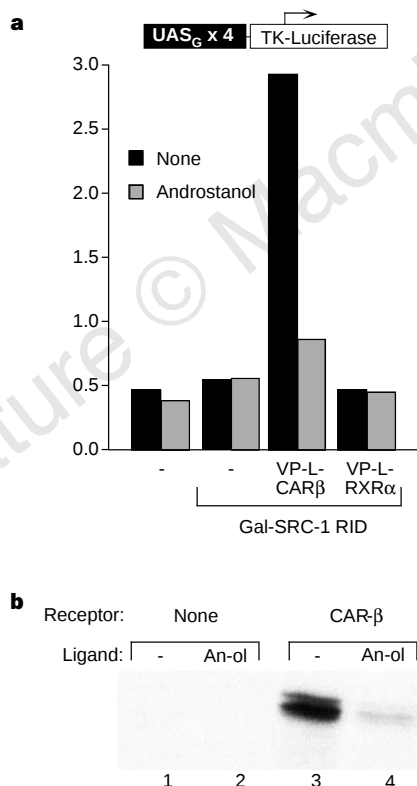


**Figure 1** CAR- $\beta$  activity is inhibited by androstane derivatives. **a**, CV-1 cells were transfected with a reporter construct containing two copies of the retinoic-acid receptor- $\beta$ 2 (RAR- $\beta$ 2) retinoic-acid-response element ( $\beta$ RAREx2-TK-luciferase), and a  $\beta$ -galactosidase expression vector, with or without expression vectors for mCAR- $\beta$ . Transfected cells were exposed to the indicated compounds (5  $\mu$ M) for 40 h and then assayed for luciferase and  $\beta$ -galactosidase activity. **b**, Chemical structures of androstanol and androstenol, the two most potent modulators of

mCAR- $\beta$ . **c**, CV-1 cells were transfected as in **a** and treated with increasing amounts of the indicated compounds. 100% of maximal activity is defined as the normalized reporter activity seen in the absence of added compounds. **d**, Left, CV-1 cells were transfected as in **c**; right, CV-1 cells were transfected with a reporter construct containing the mouse mammary tumour virus long terminal repeat (MTV-luciferase), a  $\beta$ -galactosidase expression vector and an expression vector for the human androgen receptor, and reporter activity was assayed.



**Figure 2** Inhibition by androstanol is mediated by the LBD of CAR-β. **a**, CV-1 cells were transfected with a reporter construct containing four copies of a yeast GAL4 upstream activation sequence (UAS) (MH100x4 TK-luciferase), a β-galactosidase expression vector and, where indicated, expression vectors for the human RXR-α LBD and/or a chimaeric receptor containing the LBD of mCAR-β fused to the DBD of GAL4 (GAL-mCARβ). Androstanol (5 μM) was added as indicated. **b**, Androstanol inhibits CAR-β activity in yeast. Yeast derivatives expressing LexA alone, LexA fused to the LBD of mouse CAR-β (LexA-mCARβ), or LexA-mCARβ and full-length RXR-α (ref. 26) were grown in liquid cultures in the absence or presence of 1 μM androstanol. After androstanol treatment, β-galactosidase levels were assayed using standard methods.



**Figure 3** Androstanol interacts specifically with CAR-β to promote release of an SRC-1 fragment. **a**, CV-1 cells were transfected as in Fig. 2a with the indicated constructs. GAL-SRC1 RID contains all three receptor-interaction domains of SRC-1 fused to GAL4. VP-L-mCARβ and VP-L-hRXRα contain the LBD of the indicated receptor fused to the VP16 transactivation domain. Except for cells transfected with VP-L-hRXRα, all cells were co-transfected with an expression vector for the hRXRα LBD. **b**, A fragment of SRC-1 (ref. 9) was labelled with [<sup>35</sup>S]methionine by *in vitro* translation and mixed *in vitro* with bacterially expressed GST (lanes 1, 2), or with GST fused to the full-length CAR-β (lanes 3, 4). Androstanol (An-ol, 50 μM; lanes 2, 4) was added as indicated. Bound <sup>35</sup>S-labelled SRC-1 was collected on a glutathione-Sepharose affinity column.

tanol (Fig. 3a, left panel). As expected, androstanol had no effect on VP16-RXRα, which interacted with GAL4-SRC1 only in the presence of an RXR ligand (data not shown). Similar results were obtained with GAL4-CARβ and VP16-SRC1, and by using a yeast two-hybrid system (data not shown). An *in vitro* assay of co-activator recruitment<sup>8,9</sup> showed that the interaction between CAR-β and SRC-1 is directly dissociated by the ligand androstanol (Fig. 3). Thus, compared with the glutathione S-transferase (GST) control (Fig. 3b, lanes 1, 2), bacterially expressed GST-CARβ effectively recruited a <sup>35</sup>S-labelled fragment of SRC-1 (Fig. 3b, lane 3) and this CARβ-SRC1 complex was dissociated by androstanol (Fig. 3b, lane 4). This effect was specific, as androstanol had no effect on the 9-*cis*-retinoic acid-dependent RXRα-SRC-1 complex (data not shown).

Our results define specific ligands for CAR-β and show that CAR-β acts in a manner opposite to that of conventional nuclear receptors. In classical receptors, ligand binding causes a conformational shift that places the AF-2 domain in an active conformation<sup>10,11</sup>. We predict that the CARβ LBD adopts this active conformation in the absence of ligand, and is shifted toward an inactive conformation on binding the naturally occurring inverse agonists. The androstanes are active at roughly 400 nM, which is somewhat higher than the levels of androstanol reported in the circulation of adult males<sup>12</sup>. However, this EC<sub>50</sub> is less than the level of 5α-androstans secreted by sweat<sup>13</sup> and is comparable to the levels of androstenone present in fat<sup>14</sup>. The EC<sub>50</sub> we report is also less than the micromolar EC<sub>50</sub> values reported for the ligands of several other orphan receptors<sup>6,9,15-19</sup>. Our results provide an impetus to study this previously unanticipated androstane-signalling pathway.

## Methods

**Transfections.** CV-1 (refs 20, 21), HepG2 (ref. 22) and CosM6 (ref. 23) cells were transfected as described.

**Yeast expression.** We generated derivatives of the *Saccharomyces cerevisiae* strain EGY48 (refs 24, 25) that express LexA alone, LexA fused to the LBD of mouse CAR-β (Lex-mCARβ), or LexA-mCARβ and full-length human (h)RXRα<sup>26</sup>. Yeast cells were grown in liquid cultures in the presence or absence of 1 μM androstanol, and β-galactosidase levels were assayed using standard methods.

**Assays for ligand-induced co-activator recruitment.** GST proteins were bound to glutathione-Sepharose and incubations were done in the presence of 2 mg ml<sup>-1</sup> bovine serum albumin, 25 mM HEPES (pH 7.6), 100 mM NaCl, 20% glycerol and 1 mM EDTA. Androstanol (50 μM) was added to incubations as indicated. After overnight incubation at 4°C with gentle agitation, agarose beads were extensively washed and bound proteins were eluted with 15 mM glutathione. The eluate was resolved by SDS-PAGE followed by autoradiography. The amount of <sup>35</sup>S-labelled SRC-1 bound by the GST-CARβ fusion represents ~5% of the total input.

Received 27 March; accepted 13 July 1998.

- Choi, H. S. *et al.* Differential transactivation by two isoforms of the orphan nuclear hormone receptor CAR. *J. Biol. Chem.* **272**, 23565–23571 (1997).
- Bond, R. A. *et al.* Physiological effects of inverse agonists in transgenic mice with myocardial overexpression of the β<sub>2</sub>-adrenoceptor. *Nature* **374**, 272–276 (1995).
- Klein, E. S. *et al.* Identification and functional separation of retinoic acid receptor neutral antagonists and inverse agonists. *J. Biol. Chem.* **271**, 22692–22696 (1996).
- Singer, A. G. A chemistry of mammalian pheromones. *J. Steroid Biochem. Mol. Biol.* **39**, 627–632 (1991).
- Weusten, J. J. *et al.* The mechanism of the synthesis of 16-androstenes in human testicular homogenates. *J. Steroid Biochem.* **32**, 689–694 (1989).
- Forman, B. M. *et al.* The orphan nuclear receptor LXRα is positively and negatively regulated by distinct products of mevalonate metabolism. *Proc. Natl Acad. Sci. USA* **94**, 10588–10593 (1997).
- Onate, S. A., Tsai, S. Y., Tsai, M. J. & O'Malley, B. W. Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science* **270**, 1354–1357 (1995).
- Krey, G. *et al.* Fatty acids, eicosanoids, and hypolipidemic agents identified as ligands of peroxisome proliferator-activated receptors by coactivator-dependent receptor ligand assay. *Mol. Endocrinol.* **11**, 779–791 (1997).
- Kliwer, S. A. *et al.* An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. *Cell* **92**, 73–82 (1998).
- Wagner, R. L., Apriletti, J. W., McGrath, M. E., West, B. L., Baxter, J. D. A structural role for hormone in the thyroid hormone receptor. *Nature* **378**, 690–697 (1995).
- Renaud, J. P. *et al.* Crystal structure of the RAR-γ ligand-binding domain bound to all-*trans* retinoic acid. *Nature* **378**, 681–689 (1995).

12. Bicknell, D. C. & Gower, D. B. The development and application of a radioimmunoassay for 5 $\alpha$ -androst-16-en-3 $\alpha$ -ol in plasma. *J. Steroid Biochem.* **7**, 451–455 (1976).
13. Gower, D. B. *et al.* Comparison of 16-androstene steroid concentrations in sterile apocrine sweat and axillary secretions: interconversions of 16-androstenes by the axillary microflora—a mechanism for axillary odour production in man? *J. Steroid Biochem. Mol. Biol.* **48**, 409–418 (1994).
14. Louveau, I., Bonneau, M. & Gower, D. B. Biosynthesis of 16-androstene steroids and testosterone by porcine testis tissue *in vitro*: effect of age and relationships with fat 5 $\alpha$ -androstene levels *in vivo*. *Acta Endocrinol.* **125**, 526–531 (1991).
15. Janowski, B. A. *et al.* An oxysterol signalling pathway mediated by the nuclear receptor LXR  $\alpha$ . *Nature* **383**, 728–731 (1996).
16. Lehmann, J. M. *et al.* Activation of nuclear receptor LXR by oxysterols defines a new hormone response pathway. *J. Biol. Chem.* **272**, 3137–3140 (1997).
17. Lala, D. S. *et al.* Activation of the orphan nuclear receptor steroidogenic factor 1 by oxysterols. *Proc. Natl Acad. Sci. USA* **94**, 4895–4900 (1997).
18. Forman, B. M., Chen, J. & Evans, R. M. Hypolipidemic drugs, polyunsaturated fatty acids and eicosanoids are ligands for peroxisome proliferator-activated receptors  $\alpha$  and  $\delta$ . *Proc. Natl Acad. Sci. USA* **94**, 4312–4317 (1997).
19. Kliewer, S. A. *et al.* Fatty acids and eicosanoids regulate gene expression through direct interactions with peroxisome proliferator-activated receptors  $\alpha$  and  $\gamma$ . *Proc. Natl Acad. Sci. USA* **94**, 4318–4323 (1997).
20. Forman, B. M., Umesono, K., Chen, J. & Evans, R. M. Unique response pathways are established by allosteric interactions among nuclear hormone receptors. *Cell* **81**, 541–550 (1995).
21. Forman, B. M. *et al.* Identification of a nuclear receptor that is activated by farnesol metabolites. *Cell* **81**, 687–693 (1995).
22. Seol, W., Choi, H. S. & Moore, D. D. An orphan nuclear hormone receptor that lacks a DNA binding domain and heterodimerizes with other receptors. *Science* **272**, 1136–1139 (1996).
23. Ladias, J. A. & Karathanasis, S. K. Regulation of the apolipoprotein A1 gene by ARP-1, a novel member of the steroid receptor superfamily. *Science* **251**, 561–565 (1991).
24. Golemis, E. A. & Brent, R. Fused protein domains inhibit DNA binding by LexA. *Mol. Cell. Biol.* **12**, 3006–3014 (1992).
25. Gyuris, J., Golemis, E., Charkov, H. & Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* **75**, 791–803 (1993).
26. Lee, J. W., Moore, D. D. & Heyman, R. A. A chimeric thyroid hormone receptor constitutively bound to DNA requires retinoid X receptor for hormone-dependent transcriptional activation in yeast. *Mol. Endocrinol.* **8**, 1245–1252 (1994).

**Acknowledgements.** We thank D. Russell and M. Mahendroo for discussions during this work; R. Miesfeld for an expression vector for the human androgen receptor; and S. Kliewer and I. Schulman for SRC-1 plasmids. This work was supported by the Howard Hughes Medical Institute (R.M.E.), March of Dimes (R.M.E.), CapCURE (R.M.E.), Tobacco-Related Disease Research Program (B.M.F.), The City of Hope National Medical Center/Beckman Research Institute (B.M.F.), and National Institutes of Health (D.D.M.). R.M.E. is an Investigator of the Howard Hughes Medical Institute at the Salk Institute for Biological Studies.

Correspondence and requests for materials should be addressed to B.M.F. (e-mail: bforman@gte.net).

## A Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication

Corrado Santocanale & John F. X. Diffley

Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Hertfordshire EN6 3LD, UK

DNA replication in eukaryotic cells initiates from many replication origins<sup>1</sup> which fire throughout the S phase of the cell cycle in a predictable pattern: some origins fire early, others late<sup>2</sup>. Little is known about how the initiation of DNA replication and the elongation of newly synthesized DNA strands are coordinated during S phase. Here we show that, in budding yeast, hydroxyurea, which blocks the progression of replication forks from early-firing origins, also inhibits the firing of late origins. These late origins are maintained in the initiation-competent prereplicative state for extended periods. The block to late origin firing is an active process and is defective in yeast with mutations in the *rad53* and *mec1* checkpoint genes, indicating that regulation of late origin firing may also be an important component of the 'intra-S-phase' checkpoint<sup>3</sup> and may aid cell survival under adverse conditions.

ARS305 is an efficient, early-firing replication origin<sup>4</sup>. Density-transfer experiments have shown that ARS305 fires efficiently when S phase is interrupted with the ribonucleotide reductase inhibitor hydroxyurea<sup>5</sup>. Under these conditions, forks from ARS305 stall somewhere within roughly 10 kilobases (kb) of the origin. We used two methods to examine replication intermediates generated

by these stalled forks. Either digestion of purified genomic DNA with single-strand-specific nuclease followed by neutral agarose gel electrophoresis, or denaturing agarose gel electrophoresis of undigested genomic DNA, releases replication intermediates as small fragments, whereas unreplicated genomic DNA (from the first technique) or parental DNA (from the second), which are much larger, do not run into the gel (Fig. 1a). The position of individual sequences can be visualized by hybridization with specific DNA probes. We have confirmed our major results using both assays.

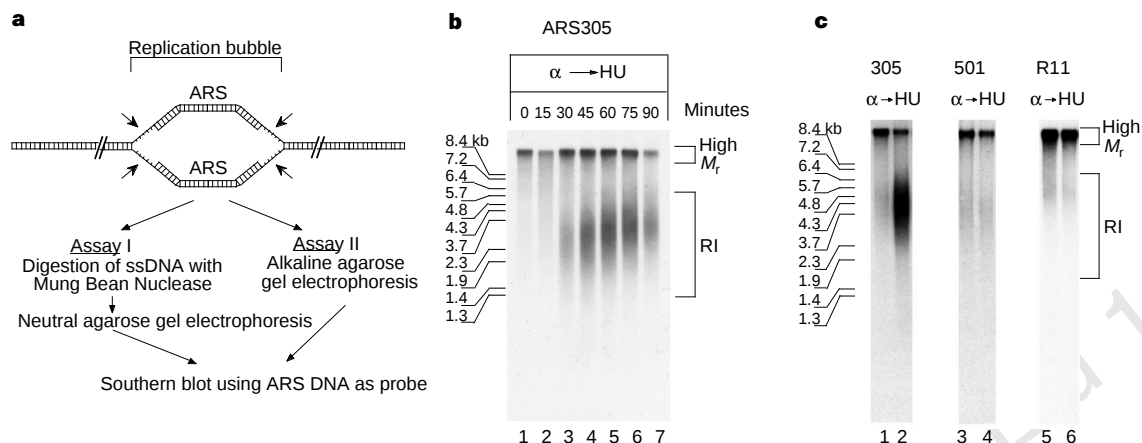
In Fig. 1b, cells were first synchronized in G1 phase by using  $\alpha$ -factor mating pheromone, and were then released into medium containing hydroxyurea. Replication intermediates from ARS305 are not present in  $\alpha$ -factor-blocked cells (Fig. 1b, lane 1); they first appear 30 min after release into hydroxyurea and persist for at least 90 min, very slowly increasing in size to an average of roughly 4–5 kb. This is consistent with the amount of replication from ARS305 seen in density-substitution experiments<sup>5</sup>. Thus, hydroxyurea does not inhibit initiation from early-firing origins but greatly slows the progression of replication forks after initiation. The accumulation of replication intermediates in hydroxyurea is origin-specific, as replication intermediates were not detected from a sequence, R11, that does not contain an origin (Fig. 1c). Most important, replication intermediates were not detected from an efficient late-firing origin, ARS501, even after long periods of time (Fig. 1c).

Cdc6p-dependent prereplicative complexes seem to be essential in the initiation of DNA replication. Genomic footprinting experiments have indicated that initiation is accompanied by loss of prereplicative complexes from origins and that new prereplicative complexes cannot assemble at origins until passage of cells through mitosis<sup>6–9</sup>. The origin ARS1, in its normal location on chromosome IV (ARS1<sup>IV</sup>), like ARS305, fires early in S phase<sup>4</sup>. The postreplicative state of ARS1<sup>IV</sup> in cells blocked in G2/M phase with nocodazole<sup>6,8</sup> is characterized primarily by the origin recognition complex (ORC)-induced DNaseI-hypersensitive site (Fig. 2a, lanes 4–6, asterisk). Lanes 7–9 show the prereplicative state of ARS1<sup>IV</sup> in cells blocked in G1 with  $\alpha$ -factor; this state is characterized by suppression of the ORC-induced hypersensitive site and protection across the B1 and B2 elements. The final three lanes (10–12) show that when cells were released from the G1 block into hydroxyurea, ARS1<sup>IV</sup> reverted to the postreplicative state. The same result was obtained for ARS305 (data not shown). This is consistent with the finding that hydroxyurea does not block the activation of these early-firing origins.

Placing ARS1 near a telomere forces it to fire later in S phase<sup>10</sup>. Therefore, we studied the effect of hydroxyurea on ARS1 in the late-firing context of a telomeric linear minichromosome (ARS1<sup>TEL</sup>). ARS1<sup>TEL</sup> forms a postreplicative complex in nocodazole-blocked cells (Fig. 2b; lanes 1, 2) and a prereplicative complex in  $\alpha$ -factor-blocked cells (Fig. 2b, lanes 3, 4); these complexes are indistinguishable from those formed at ARS1<sup>IV</sup>. In contrast to ARS1<sup>IV</sup>, however, ARS1<sup>TEL</sup> remains prereplicative after release into hydroxyurea for at least 90 min (compare Fig. 2a and b). Therefore, conversion of ARS1 to a late-firing origin renders its firing sensitive to inhibition by hydroxyurea. Consequently, the inhibition of origin firing by hydroxyurea does not appear to be an intrinsic property of specific origin sequences but is dictated by their time of firing during S phase.

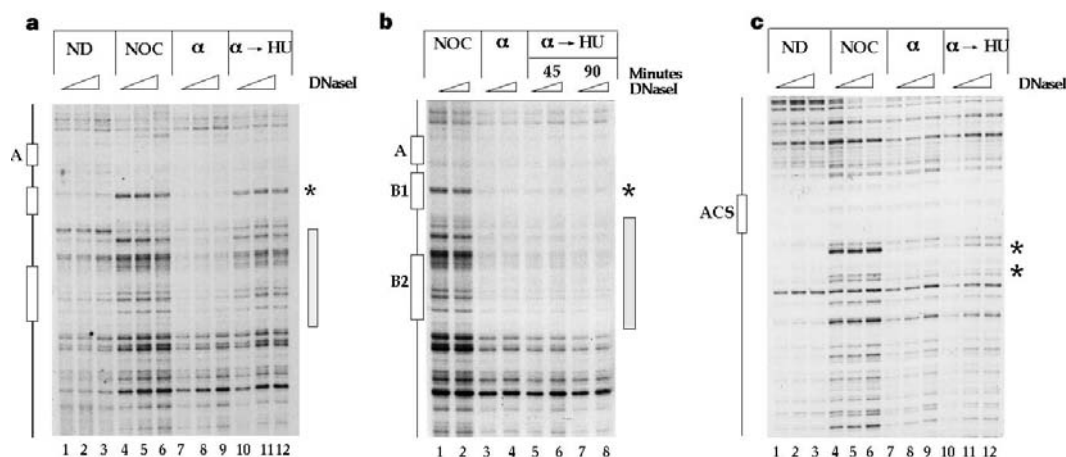
ARS301 is inactive in its normal chromosomal location but active >95% of the time when present on a plasmid<sup>8,11,12</sup>. On a plasmid, ARS301 is activated after ARS305 (ref. 5). Density-substitution experiments have shown that there is no detectable DNA replication from this plasmid in hydroxyurea<sup>5</sup>. When we used the DNA samples from Fig. 2a to study the status of plasmid-borne ARS301, we found that, unlike ARS1<sup>IV</sup> and ARS305, but like ARS1<sup>TEL</sup>, ARS301 remained prereplicative after release into hydroxyurea as shown by continued suppression of the ORC-induced hypersensitive sites (asterisks, Fig. 2c). Plasmid-borne ARS301 reverts to the postreplicative complex, and the plasmid duplicates in a Cdc7-depen-





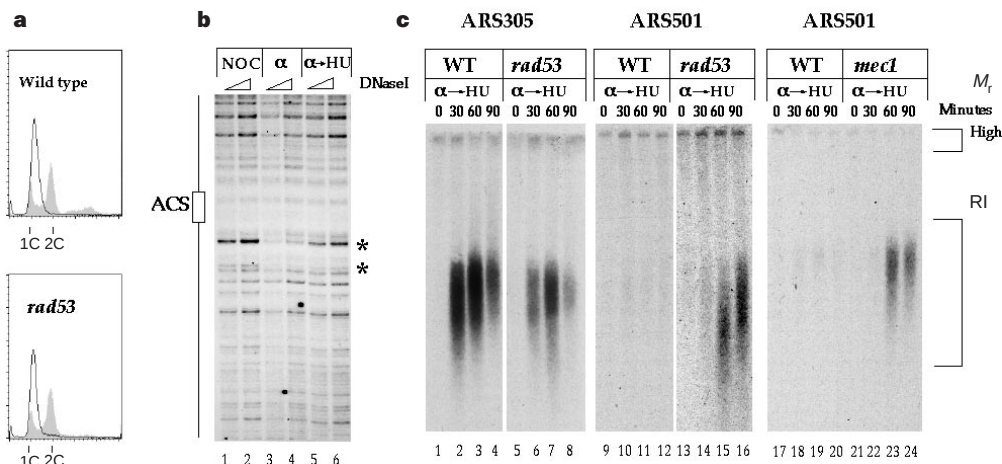
**Figure 1** Analysis of replication intermediates (RIs) from stalled replication forks. **a**, The assays used to detect RIs. Single-stranded (ss) DNA is indicated by the short arrows. ARS, origin sequence. **b**, Detection of RIs from chromosomal ARS305.  $\alpha$ -factor-arrested cells (Y300 (ref. 23)) were released into medium containing hydroxyurea (HU) for 0–90 min. Replication intermediates from

ARS305 were studied using mung bean nuclease (see Methods). **c**, RI accumulation is specific for an early-firing origin. At 0 and 90 min after release of cells from  $\alpha$ -factor, samples were collected and processed as above. Probes for RI detection were specific for the early-firing origin ARS305 (305), the late-firing origin ARS501 (501) and the non-origin-containing sequence R11.



**Figure 2** Conversion from the pre- to the postreplicative state is blocked at late firing origins of replication in hydroxyurea. **a**, Genomic footprinting analysis of ARS1<sup>IV</sup>. Cells (W303-1a) carrying plasmid pCS1 (ref. 8) were arrested either in G2 phase with nocodazole (NOC) or in G1 phase with  $\alpha$ -factor ( $\alpha$ ), or were released from  $\alpha$ -factor block into hydroxyurea for 90 min ( $\alpha \rightarrow$  HU). ND, naked DNA. Asterisk indicates ORC-induced hypersensitive sites and grey bar indicates G1-

specific protection. **b**, ARS1<sup>TEL</sup> firing is sensitive to inhibition by hydroxyurea. Cells (W303-1a,  $\Delta$ ARS1<sup>IV</sup>:HIS3) carrying the linear telomeric plasmid pYLPV (ref. 24) were treated as in **a**. Chromatin structure of plasmid-borne ARS1<sup>IV</sup> could be unambiguously detected because ARS1<sup>IV</sup> was deleted in this strain. **c**, Activation of plasmid-borne ARS301 is blocked by hydroxyurea. DNA samples from **a** were used to analyse the chromatin structure of ARS301 on plasmid pCS1.



**Figure 3** Inhibition of late origin firing is an active process. **a**, Bulk DNA replication is blocked by hydroxyurea in both wild-type and *rad53* mutant strains. DNA content of wild-type (Y300 (ref. 23)) and *rad53* mutant (Y301 (ref. 23)) cells released from G1 into hydroxyurea-containing medium for 90 min (black lines) was measured by flow cytometry. FACS profiles from logarithmically growing cells (grey areas) are shown to indicate the positions of the 1C and 2C peaks. **b**, Plasmid-

borne ARS301 loses its prereplicative complex in *rad53* mutant cells in hydroxyurea *rad53* mutant cells carrying plasmid pCS1 were treated as in Fig. 2c. **c**, ARS501 fires in hydroxyurea in checkpoint-deficient strains.  $\alpha$ -factor-arrested cells (Y300, Y301, YMP10860, YMP10848 (ref. 25)) were released into hydroxyurea-containing medium. RIs from ARS305 and ARS501 were studied using the alkaline gel electrophoresis method.

dent reaction, after release from hydroxyurea<sup>5</sup>, showing that ARS301 remains competent to fire during the hydroxyurea block. These results show that firing of late origins is blocked in hydroxyurea at the conversion from the pre- to the postreplicative complex.

Hydroxyurea activates a checkpoint that delays entry into mitosis until S phase is completed<sup>13</sup>. We therefore studied known checkpoint mutants for defects in the dependency of late origin firing on early replication. *rad9*, *mec3* or *tell* deletion mutants were all still capable of blocking ARS301 activation in hydroxyurea (data not shown); however, the *mec1-1* mutant and two different *rad53* mutant alleles (also known as *sad1*, *mec2* and *spk1*) were defective in this dependency (Fig. 3; data not shown). After release from the  $\alpha$ -factor block into hydroxyurea, *rad53* mutant cells, like wild-type cells, do not replicate significant amounts of DNA, arresting with a near 1C DNA content (Fig. 3a). Thus, the block to bulk DNA synthesis by hydroxyurea does not depend on Rad53. However, unlike wild-type cells, ARS301 is converted to the postreplicative state in *rad53* (compare Fig. 2c and Fig. 3b) and *mec1* (data not shown) mutants after release into hydroxyurea. This is because late origins fire inappropriately in these checkpoint mutants (Fig. 3c). Replication intermediates from ARS305 can be detected in both wild-type cells and *rad53* mutants blocked in hydroxyurea (Fig. 3c, lanes 1–8), indicating that ARS305 is activated and that replication forks stall in both strains. However, replication intermediates from ARS501 are not detected in wild-type cells blocked in hydroxyurea (Fig. 1c, lanes 9–12 and Fig. 3c, lanes 9–12, 17–20) whereas they can clearly be detected from ARS501 in either *rad53* (lanes 13–16) or *mec1* (lanes 21–24) mutant cells blocked in hydroxyurea. Replication intermediates were not detected from non-origin-containing sequences such as R11 in these experiments (Supplementary information), indicating that accumulation of replication intermediates is origin-specific even in the checkpoint mutants. These results show that Rad53 and Mec1 are required to block the activation of late-firing origins in hydroxyurea.

ARS501 replication intermediates appear 30 min later than ARS305 in both mutants (Fig. 3c), indicating that the relative order of origin firing is retained. We have also found that ARS501 still fires after ARS305 in *rad53* mutants in an unperturbed S phase, although its firing was slightly accelerated relative to ARS305 (Supplementary information). It is likely that establishment of

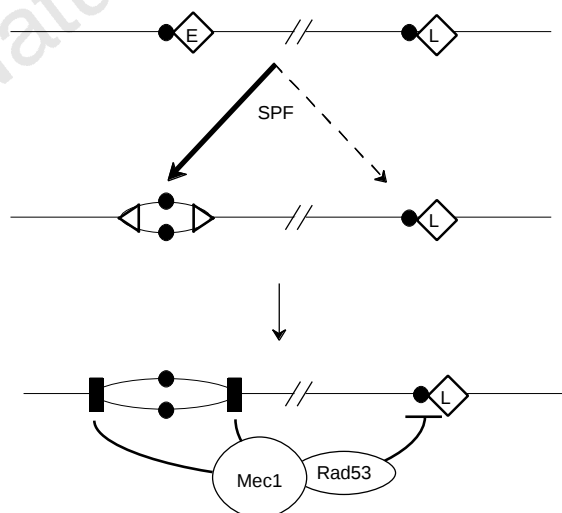
replication forks from early-firing origins is required so that hydroxyurea can generate the signals that activate the checkpoint pathway. Thus the Mec1/Rad53 pathway is unlikely to be solely responsible for establishing the normal order of origin firing.

We propose the model shown in Fig. 4. During G1 phase, prereplicative complexes assemble at future replication origins. This is roughly the time at which origins become committed to either early or late firing in the subsequent S phase<sup>14</sup>. Early origins are preferentially activated by S-phase-promoting factors, such as the protein kinases Cdc28 and Cdc7, by an unknown mechanism. If replication forks stall, the firing of late origins is blocked by the Mec1/Rad53 protein kinases. Given the homology between Mec1 and the human ataxia telangiectasia mutated (ATM) protein<sup>15</sup>, we suggest that the regulation of late origin firing described here may be related to the fact that  $\gamma$ -irradiation blocks activation of new replicons in mammalian cells at an undefined step and that this block is defective in ATM mutant cell lines<sup>16,17</sup>. We also suggest that the Mec1/Rad53-dependent inhibition of late origin firing may help to explain the expansion of S phase when DNA has been damaged by methylating agents<sup>3</sup>. Further characterization of this 'origin-firing checkpoint' should help in determining how S-phase events are coordinated in eukaryotic cells and how anti-cancer drugs that target DNA replication, such as hydroxyurea, work. □

## Methods

Cell-cycle arrests, fluorescence-activated cell sorting (FACS) and genomic footprinting were performed as described<sup>16,18–20</sup>. Hydroxyurea was used at 0.2 M at 25 °C in all experiments. For the analysis of replication intermediates, genomic DNA was prepared as described<sup>2</sup>. DNA prepared from 10<sup>7</sup> cells was digested for 1 h at 30 °C with 10 U of mung bean nuclease (New England Biolabs) and separated on 0.8% neutral gels in Tris-acetate-EDTA buffer<sup>21</sup>. Alkaline gel electrophoresis was performed using 1% agarose gels as described<sup>21</sup>. DNA was transferred from gels to Hybond N<sup>+</sup> membranes according to the manufacturer's instructions. Probes for ARS305 origin and R11 sequence on chromosome V have been described<sup>4,22</sup>. A probe specific for the late-replicating origin ARS501 was amplified by the polymerase chain reaction from genomic DNA using the oligonucleotides 5'-GGA TCC CGA GTC ATG TTT GG-3' and 5'-GAG CAT AAT TAT GAC TGT AGC CC-3'.

Received 16 March; accepted 15 July 1995.



**Figure 4** Model for regulation of late-firing origins. During G1 phase, prereplicative complexes (diamonds) assemble at both early (E)- and late (L)-firing origins of replication. Early origins are preferentially activated by S-phase-promoting factors (SPF). If replication forks (triangles) stall (black rectangles), the firing of late origins is blocked by the activity of Mec1 and Rad53. Arrows from SPF indicate the relative activity of SPF towards early and late origins.

- Huberman, J. A. & Riggs, A. D. On the mechanism of DNA replication in mammalian chromosomes. *J. Mol. Biol.* **32**, 327–341 (1968).
- Fangman, W. L. & Brewer, B. J. A question of time—replication origins of eukaryotic chromosomes. *Cell* **71**, 363–366 (1992).
- Paulovich, A. G. & Hartwell, L. H. A checkpoint regulates the rate of progression through S phase in *S. cerevisiae* in response to DNA damage. *Cell* **82**, 841–847 (1995).
- Reynolds, A. E., McCarroll, R. M., Newlon, C. S. & Fangman, W. L. Time of replication of ARS elements along yeast chromosome III. *Mol. Cell. Biol.* **9**, 4488–4494 (1989).
- Bousset, K. & Diffley, J. F. X. The Cdc7 protein kinase is required for origin firing during S phase. *Genes Dev.* **12**, 480–490 (1998).
- Diffley, J. F. X., Cocker, J. H., Dowell, S. J. & Rowley, A. Two steps in the assembly of complexes at yeast replication origins *in vivo*. *Cell* **78**, 303–316 (1994).
- Cocker, J. H., Piatti, S., Santocanele, C., Nasmyth, K. & Diffley, J. F. X. An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. *Nature* **379**, 180–182 (1996).
- Santocanele, C. & Diffley, J. F. X. ORC- and Cdc6-dependent complexes at active and inactive chromosomal replication origins in *Saccharomyces cerevisiae*. *EMBO J.* **15**, 6671–679 (1996).
- Piatti, S., Bohm, T., Cocker, J. H., Diffley, J. F. X. & Nasmyth, K. Activation of S-phase promoting CDKs in late G1 defines a “point of no return” after which Cdc6 synthesis cannot promote DNA replication in yeast. *Genes Dev.* **10**, 1516–1531 (1996).
- Ferguson, B. M. & Fangman, W. L. A position effect on the time of replication origin activation in yeast. *Cell* **68**, 333–339 (1992).
- Dubey, D. D. *et al.* Evidence suggesting that the ARS elements associated with silencers of the yeast mating-type locus HML do not function as chromosomal DNA replication origins. *Mol. Cell. Biol.* **11**, 5346–5355 (1991).
- Broach, J. R. *et al.* Localization and sequence analysis of yeast origins of DNA replication. *Cold Spring Harb. Symp. Quant. Biol.* **47**, 1165–1173 (1982).
- Elledge, S. J. Cell cycle checkpoints: preventing an identity crisis. *Science* **274**, 1664–1672 (1996).
- Raghuraman, M. K., Brewer, B. J. & Fangman, W. L. Cell cycle-dependent establishment of a late replication program. *Science* **276**, 806–809 (1997).
- Zakian, V. A. ATM-related genes: what do they tell us about functions of the human gene? *Cell* **82**, 685–687 (1995).
- Painter, R. B. & Young, B. R. Radiosensitivity in ataxia-telangiectasia: a new explanation. *Proc. Natl. Acad. Sci. USA* **77**, 7315–7317 (1980).
- Larner, J. M., Lee, H. & Hamlin, J. L. Radiation effects on DNA synthesis in a defined chromosomal replicon. *Mol. Cell. Biol.* **14**, 1901–1908 (1994).
- Santocanele, C., Neeke, H., Longhese, M. P., Lucchini, G. & Plevani, P. Mutations in the gene encoding the 34 kDa subunit of yeast replication protein A cause defective S phase progression. *J. Mol. Biol.* **254**, 595–607 (1995).

19. Rowley, A., Cocker, J. H., Harwood, J. & Diffley, J. F. X. Initiation complex assembly at budding yeast replication origins begins with the recognition of a bipartite sequence by limiting amounts of the initiator, ORC. *EMBO J.* **14**, 2631–2641 (1995).
20. Diffley, J. F. X. & Cocker, J. H. Protein–DNA interactions at a yeast replication origin. *Nature* **357**, 169–172 (1992).
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harb. Lab. Press, Cold Spring Harbor, 1989).
22. Ferguson, B. M., Brewer, B. J., Reynolds, A. E. & Fangman, W. L. A yeast origin of replication is activated late in S phase. *Cell* **65**, 507–515 (1991).
23. Allen, J. B., Zhou, Z., Siede, W., Friedberg, E. C. & Elledge, S. J. The SAD1/RAD53 protein kinase controls multiple checkpoints and DNA damage-induced transcription in yeast. *Genes Dev.* **8**, 2401–2415 (1994).
24. Wellinger, R. J., Ethier, K., Labrecque, P. & Zakian, V. A. Evidence for a new step in telomere maintenance. *Cell* **85**, 423–433 (1996).
25. Paulovich, A. G., Margulies, R. U., Garvik, B. M. & Hartwell, L. H. RAD9, RAD17, and RAD24 are required for S phase regulation in *Saccharomyces cerevisiae* in response to DNA damage. *Genetics* **145**, 45–62 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

**Acknowledgements.** We thank S. Elledge and L. Hartwell for yeast strains; V. A. Zakian for plasmid pYLPV; members of the Diffley lab for stimulating discussions; and members of the ICRF Photography, Peptide Synthesis and Oligonucleotide Synthesis departments for their help.

Correspondence and requests for materials should be addressed to J.F.X.D. (J. Diffley@icrf.icnet.uk).

## Regulation of DNA-replication origins during cell-cycle progression

Katsuhiko Shirahige\*, Yuji Hori\*, Katsuya Shiraishi\*, Minoru Yamashita†, Keiko Takahashi\*, Chikashi Obuse\*, Toshiaki Tsurimoto\* & Hiroshi Yoshikawa\*

\* Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, Japan

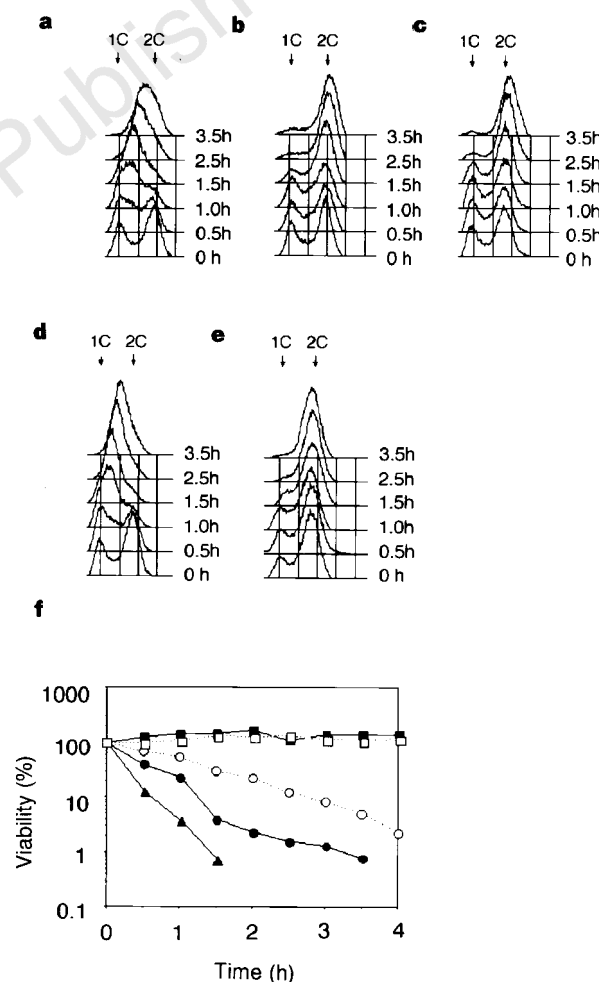
† Ajinomoto Co. Inc., 1-1 Suzuki, Kawasaki, Kawasaki City 210-8680, Japan

We have shown previously that chromosome VI of *Saccharomyces cerevisiae* contains nine origins of DNA replication that differ in initiation frequency and replicate sequentially during the S phase of the cell cycle<sup>1,2</sup>. Here we show that there are links between activation of these multiple origins and regulation of S-phase progression. We study the effects of a DNA-damaging agent, methyl methane sulphonate (MMS), and of mutations in checkpoint genes such as *rad53* (ref. 3) on the activity of origins, measured by two-dimensional gel analysis, and on cell-cycle progression, measured by fluorescence-activated cell sorting. We find that when MMS slows down S-phase progression it also selectively blocks initiation from late origins. A *rad53* mutation enhances late and/or inefficient origins and releases the initiation block by MMS. Mutation of *rad53* also results in a late origin becoming early replicating. We conclude that *rad53* regulates the timing of initiation of replication from late origins during normal cell growth and blocks initiation from late origins in MMS-treated cells. *rad53* is, therefore, involved in the cell's surveillance of S-phase progression<sup>4,5</sup>. We also find that *orc2*, which encodes subunit 2 of the origin-recognition complex<sup>6,7</sup>, is involved in suppression of late origins.

Among four origins on the right arm of chromosome VI, ori607 initiated replication immediately early in S phase and a telomere-proximal origin, ori609, initiated 30 min later. Two other origins, ori606 and ori608 (an inefficient origin used in only 5% of the cell cycle), initiated 5–10 min later than ori607. Replication of the 100-kilobase (kb) telomere-proximal region of the right arm of chromosome VI was unusually slow and varied among different cells of a population. We thought that such a slow and heterogeneous rate of replication of this region of the chromosome might be due to the downregulation of replication of this region in response to physiological variations of individual cells. To test this hypothesis, we

studied the effect of MMS on the replication of the four origin regions on the right arm of chromosome VI, because damage to DNA in general blocks S-phase progression<sup>3,8</sup>. We also studied the effects of *rad53* and *mec1* mutations because the products of these genes are involved in the cell's surveillance of S-phase progression<sup>3,4,8</sup>.

First, we studied the effect of MMS on cell-cycle progression and cell viability. In the wild-type cell, MMS blocked cell-cycle progression during S phase without affecting viability, indicating that the cell's surveillance mechanism is functioning in S phase in MMS-treated cells (Fig. 1a, d, f). In contrast, cells continued to proceed through S phase to G2 phase concomitantly with the loss of viability in *rad53-1* and *mec1-1* mutant cells (Fig. 1b, c, f). We also studied *orc2-1* mutant cells; the viability of these cells was sensitive to MMS at the permissive temperature (Fig. 1f) and the cells were not blocked by MMS during S-phase progression (Fig. 1d, e). To confirm these results we studied the effect of MMS on the progression through S phase of cells that had been arrested at G1 phase



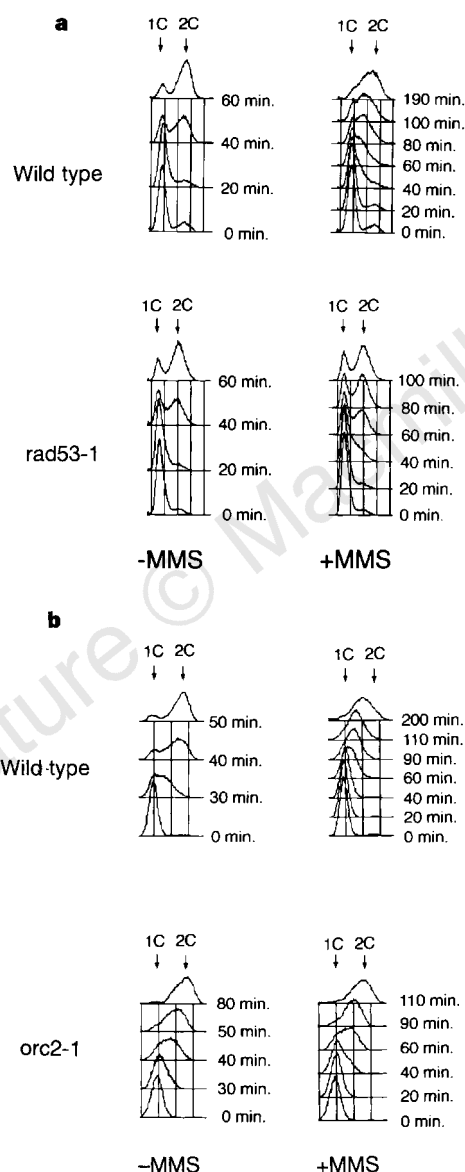
**Figure 1** Effect of the *rad53-1*, *mec1-1* and *orc2-1* mutations on cell-cycle progression and viability of MMS-treated cells. Cells were grown in YPDA medium to  $A_{600} = 0.6$  (0 h) and MMS (0.015%) was added. **a–e**, Cell-cycle distribution and **f**, cell viability were measured by FACS (**a–e**) and plating on YPDA (**f**) at the indicated times after MMS addition. **a–e**, 1C and 2C indicate cells with one and two sets of chromosomes, respectively. **a**, Wild-type strain 7830-2-4A. **b, c**, Its two mutants, *rad53-1* (strain DLY264) (**b**) and *mec1-1* (strain DLY285) (**c**), all grown at 30°C. **d**, Wild-type strain W303-1B. **e**, Its *orc2-1* mutant strain (JRY4125) both grown at 26°C. **f**, Open squares, strain 7830-2-4A; filled squares, strain W303-1B; filled circles, *rad53-1* mutant; filled triangle, *mec1-1* mutant; open circles, *orc2-1* mutant.

by  $\alpha$ -factor to synchronize their cell cycles (Fig. 2a). The parental cells showed a dramatic reduction in S-phase progression in the presence of MMS, whereas S-phase progression of *rad53-1* cells was slower than normal but proceeded to completion in the presence of MMS. This is a well-known phenomenon which is ascribed to checkpoint regulation by *rad53* (ref. 3). The effect of the *orc2-1* mutation on S-phase progression in the presence of MMS was less clear, but indicated that *orc2* may also be involved in the checkpoint regulation of S-phase progression (Fig. 2b).

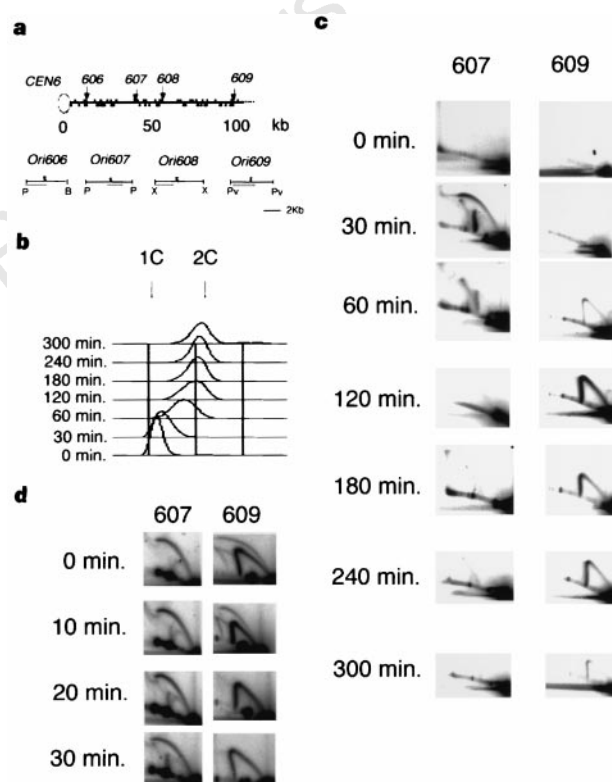
To understand how MMS blocks S-phase progression, we investigated the effect of MMS on the activation of two origins of the right arm of the chromosome VI in wild-type cells (Fig. 3a, b). In the presence of MMS, the initiation of replication from the early origin, ori607, occurred at ~30 min after release from G1 block, whereas

the late origin, ori609, did not fire during the time studied (Fig. 3c). This is consistent with the fact that the cell did not complete S phase in the presence of MMS (Fig. 3b). Passive replication of the late-origin region started at ~60 min but proceeded slowly; Y-fork intermediates accumulated for an extended period of time (Fig. 3c). These results indicate that MMS does not affect the initiation of the replication from the early origin, whereas it blocked initiation from late origins and inhibited passive replication of the late-origin region. We observed the same selective inhibition of initiation between early and late origins in exponentially growing cells (Fig. 3d).

As this is a simple and sensitive method with which to study the selective response of early and late origins to MMS, we determined the effect of MMS on the replication of the four origins on the right



**Figure 2** Effect of the *rad53-1* and *orc2-1* mutations on S-phase progression of MMS-treated cells. **a**, **b**, *rad53-1* (strain DLY264) and *orc2-1* (strain YHY201A) mutants and their parental strains (7830-2-4A and W303-1A, respectively) were synchronized in G1 by  $\alpha$ -factor (6  $\mu$ M) as described<sup>1,3</sup> and then released to S phase in the presence or absence of 0.03% MMS. Cells were grown **a**, at 30°C or **b**, at 26°C. At the times indicated, samples were taken for FACS analysis. Positions of 1C and 2C cells were determined by FACS analysis of a random culture of each strain.



**Figure 3** Effect of MMS on initiation of replication from early and late origins on chromosome VI. **a**, Top, the map of *Hind*III (above the line) and *Eco*RI (below the line) restriction sites is shown together with the location of the origins ori606–609 on the right arm of chromosome VI. Bottom, the physical map of the fragments ori606 (5.7 kb), 607 (5.9 kb), 608 (7.1 kb) and 609 (6.1 kb) used for two-dimensional gel analysis, together with the positions of the probes used (dotted lines). **B**, *Bam*HI; *E*, *Eco*RI; *H*, *Hind*III; *P*, *Pst*I; *Pv*, *Pvu*II; *X*, *Xho*I. **b**, **c**, 7830-2-4A cells were synchronized in G1 by  $\alpha$ -factor (6  $\mu$ M), and incubated with 0.015% MMS for 30 min before the release to S phase in the presence of 0.015% MMS at 23°C. At the indicated times samples were taken for FACS<sup>3</sup> (**b**) and two-dimensional gel<sup>112</sup> (**c**) analyses. Positions of 1C and 2C cells are as in Fig. 1. **d**, 7830-2-4A cells were grown to  $A_{600} = 0.6$  and MMS (0.015%) was added. The culture was taken at indicated times after the MMS addition for two-dimensional gel analysis. The fragments shown in **a** were used for two-dimensional gel analysis in **c**, **d**.

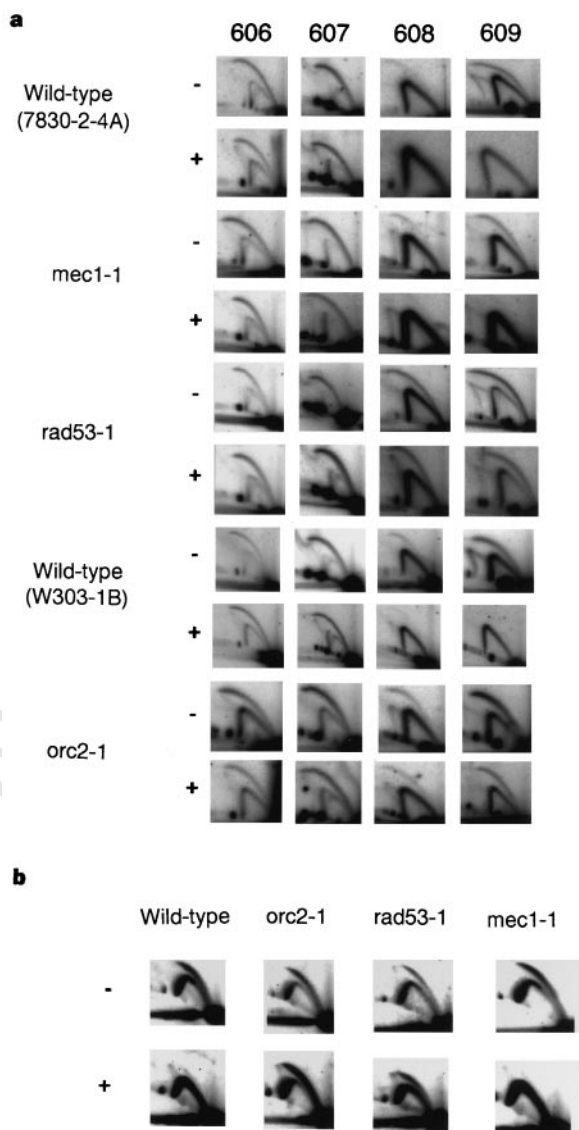


arm of chromosome VI under the influence of *rad53*, *mec1* and *orc2* mutations (Fig. 4a). In the parental cells, MMS blocked the initiation from the late-replicating and inefficient origins (ori609 and ori608, respectively) but did not affect initiation from immediate-early and early origins (ori607 and ori606, respectively). In contrast, we did not detect a block of initiation of late and inefficient origins in either *rad53-1* or *orc2-1* mutant cells. These two mutations enhanced the initiation frequency from the late and/or weak origins in normal growth conditions in the absence of MMS. Unexpectedly, MMS did not cause the selective inhibition of initiation from the late origin in a *mec1-1* mutant cell that exhibited sensitivity to MMS (Fig. 1f) and did not block S-phase progression similarly to *rad53-1* (Fig. 1c).

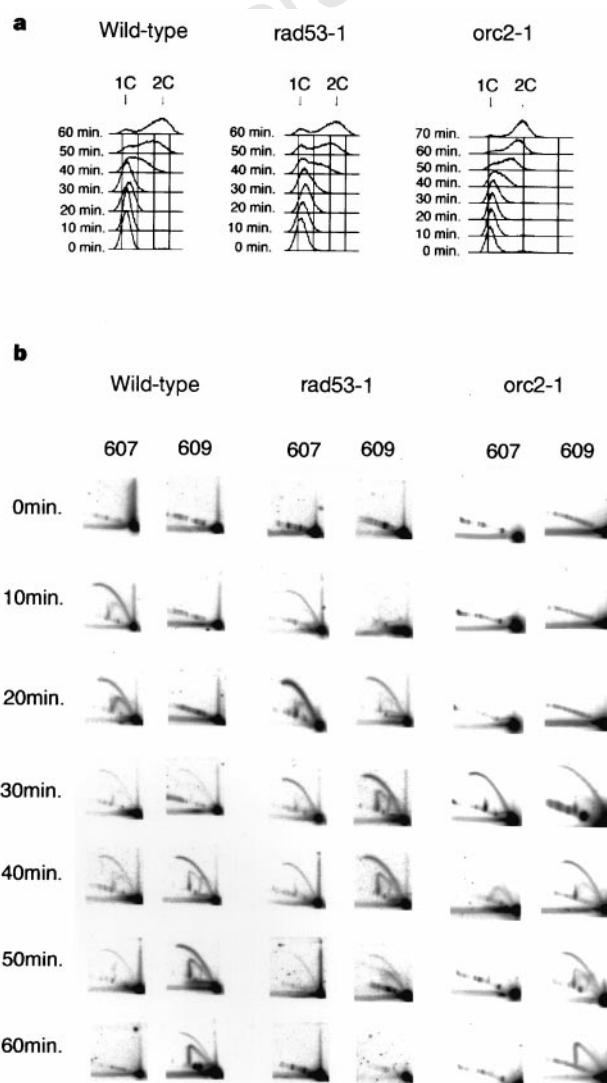
To confirm the generality of these results, we studied another efficient, late-firing origin, ori501, which is located on chromosome V (Fig. 4b)<sup>9</sup>. MMS completely blocked initiation from ori501 in

wild-type and *mec1-1* cells but did not block initiation of this origin in *rad53-1* and *orc2-1* cells. These results show that Rad53 and Orc2 are involved in the suppression of the late and weak origins, and indicate that the activation of *rad53* blocks completely the initiation from these origins in response to DNA damage by MMS. The function of *orc2* in the suppression is unknown. In contrast to *rad53-1*, the *mec1-1* mutation did not affect initiation of late and inefficient origins in both the presence and the absence of MMS. This may be due to the presence of a suppressor mutation, *sm11*, in the *mec1-1* mutant strain<sup>8</sup>.

We then studied the effect of *rad53* and *orc2* mutations on the timing of initiation and passive replication of ori607 and ori609 to determine whether these genes are involved in controlling the timing of activation of the early and late origins (Fig. 5a, b). During synchronous S-phase progression of wild-type cells, ori607 fired 10 min after the release to S phase, before the start of



**Figure 4** Effect of the *rad53-1*, *mec1-1* and *orc2-1* mutations on the activity of several replication origins of the cell in the absence or presence of MMS. The three mutant strains and their parental strains are as in Fig. 1. Cells were grown with or without MMS. Samples were taken at 30 min after MMS addition and were analysed as in Fig. 3d. **a**, Four origins (ori606–609) on the right arm of the chromosome VI were analysed as described in Fig. 3. **b**, We performed the same analysis for ori501 of chromosome V. A 3.2-kb *Xba*I-digested chromosomal fragment containing ori501 was probed by an ori501-specific probe as described<sup>9</sup>.



**Figure 5** Effect of the *rad53-1* and *orc2-1* mutations on the timing of initiation and passive replication of two origins of chromosome VI. Growth and synchronization of the two mutants and their parental cells were as in Fig. 2. Cells were then released into S phase in YPDA medium at 23 °C (time 0). **a**, FACS analysis and **b**, two-dimensional gel analysis of two origins are as described in Fig. 3b. Positions of 1C and 2C cells were determined from the pattern of the exponential culture run side by side.

DNA synthesis is detectable by fluorescence-activated cell sorting (FACS) (Fig. 5a), whereas the initiation of ori609 started at 30 min and continued until 60 min after the release to the S phase. In *rad53-1* mutant cells the early origin fired at the same time as in the wild type, whereas the late origin fired more than 10 min earlier than in the wild-type cells and continued for 30 min. In addition, the characteristic long time span of passive replication of the ori609 region in wild-type cells disappeared completely in the *rad53-1* mutant cells. Thus the timing of firing of the late origin and the regulation of replication-fork movement at the terminal region of the right half of the chromosome VI (ref. 1) are disturbed by *rad53-1* mutation. The *orc2-1* mutation affected the timing of firing of both early and late origins without affecting fork movement in the late-origin region.

Mutations in other genes involved in initiation and elongation of DNA replication (*cdc6-1*, *mcm3-1*, *pol2-11*, *dpb11-1*, *cdc47-1* and *dna2* mutants) had no effect on the block of S-phase progression although these mutants exhibited increased sensitivity to MMS. We were surprised to find that Rad53 and Orc2 are involved in regulating the efficiency of initiation of origins in normal growth conditions. In fact, a very inefficient origin on chromosome VI, ori608, was activated quite efficiently in both *rad53-1* and *orc2-1* mutants (Fig. 4a). Moreover, the timing of initiation of the late origin ori609 was markedly accelerated in the *rad53-1* mutant, and ori609 became an earlier-initiating origin (Fig. 5b). These results indicate that *rad53* is involved in determining the efficiency and timing of late-replicating origins of chromosomes by suppressing the activation of these origins. *orc2* may be involved in the regulation of activation of early origins and in suppression of late-origin firing. In both cases, the suppression of late origins is important in the cell's surveillance of S-phase progression when cells are exposed to DNA-damaging reagents.

In mammalian cells, the reduction in S-phase progression in the presence of DNA damage may be the result of inactivation of some origins<sup>10</sup>. For example, in CHO cells, ionizing radiation caused the inhibition of initiation of origins at the *dihydrofolate reductase* locus rather than inhibition of elongation<sup>10,11</sup>, indicating the same downregulation system may be conserved in yeast and higher eukaryotes. □

## Methods

We used the following yeast strains: W303-1A (genotype MAT-a *ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1*), W303-1B (genotype MAT-α *ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1*, both from R. Rothstein), 7830-2-4A (genotype MAT-a *his3 leu2 ura3 trp1*), DLY264 (genotype MATa *rad53-1* (7830-2-4A background), DLY285 (genotype MATa *mec1-1 sml1* (7830-2-4A background), all three from L. H. Hartwell), JRY4125 (genotype MAT-α *orc2-1* (W303 background), from J. Rine), YHY201A (genotype MAT-a *orc2-1* (W303 background)), and YHY301A (genotype MAT-a *rad53-1* (W303 background)).

Cell-cycle arrest and two-dimensional gel-electrophoresis analyses of replication intermediates have been described<sup>2,12</sup>. Initiation and replication time analyses were performed as described<sup>2</sup>. FACS analyses were done as described<sup>3</sup>.

Received 16 July; accepted 7 August 1998.

1. Friedman, K. L., Brewer, B. J. & Fangman, W. L. Replication profile of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* **2**, 667–678 (1997).
2. Yamashita, M. et al. The efficiency and timing of initiation of replication of multiple replicons of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* **2**, 655–665 (1997).
3. Paulovich, A. G. & Hartwell, L. H. A checkpoint regulates the rate of progression through S phase in *S. cerevisiae* in response to DNA damage. *Cell* **82**, 841–847 (1995).
4. Elledge, S. J. Cell cycle checkpoints: preventing an identity crisis. *Science* **274**, 1664–1672 (1996).
5. Nasmyth, K. Viewpoint: putting the cell cycle in order. *Science* **274**, 1643–1645 (1996).
6. Foss, M., McNally, F. J., Laurenson, P. & Rine, J. Origin recognition complexes (ORC) in transcriptional silencing and DNA replication in *S. cerevisiae*. *Science* **262**, 1838–1844 (1993).
7. Bell, S. P., Kobayashi, R. & Stillman, B. Yeast origin recognition complex functions in transcription silencing and DNA replication. *Science* **262**, 1844–1849 (1993).
8. Paulovich, A. G., Margulies, R. U., Garvik, B. M. & Hartwell, L. H. *RAD9*, *RAD17*, and *RAD24* are required for S phase regulation in *Saccharomyces cerevisiae* in response to DNA damage. *Genetics* **145**, 45–62 (1997).
9. Ferguson, B. M., Brewer, B. J., Reynolds, A. E. & Fangman, W. L. A yeast origin of replication is activated late in S phase. *Cell* **65**, 507–515 (1991).
10. Larner, J. M., Lee, H. & Hamlin, J. L. Radiation effects on DNA synthesis in a defined chromosomal replicon. *Mol. Cell. Biol.* **14**, 1901–1908 (1994).
11. Lee, H., Laner, J. M. & Hamlin, J. L. A p53-independent damage-sensing mechanism that functions as a checkpoint at the G1/S transition in Chinese hamster ovary cells. *Proc. Natl Acad. Sci. USA* **94**, 526–531 (1997).
12. Hori, Y., Shirahige, K., Obuse, C., Tsurimoto, T. & Yoshikawa, H. Characterization of a novel CDC gene (*ORC1*) partly homologous to *CDC6* of *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **7**, 409–418 (1996).

**Acknowledgements.** We thank J. Rine and B. Stillman for the *orc2-1* mutant and its parental strain; L. H. Hartwell for the *rad53-1* mutant, the *mec1-1* mutant and their parental strain (7830-2-4A); and J. F. X. Diffley for critical reading of this manuscript and for unpublished results. This work was supported by grants-in-aid for Cooperative Research and for Special Projects Research from the Ministry of Education, Science and Culture, Japan, and by Special Coordination Funds for promoting Science and Technology from the Science and Technology Agency of the Japanese Government.

Correspondence and requests for materials should be addressed to H.Y. (e-mail: hyoshika@bs.aist-nara.ac.jp).

## erratum

# Optical alignment and spinning of laser-trapped microscopic particles

M. E. J. Friese, T. A. Nieminen, N. R. Heckenberg & H. Rubinsztein-Dunlop

*Nature* **394**, 348–350 (1998)

The three sequential frames that should have been depicted in Fig. 1 were in fact all identical. The correct sequence is shown here. □

